

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 13:47:50 ; Search time 2199.24 Seconds  
(without alignments)  
1283.613 Million cell updates/sec

Title: US-09-854-133-586

Perfect score: 532  
Sequence: 1 EVEYSRDHSLGDSLTLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_173955\_334/app\_query.fasta\_1.462  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=oct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133.cgn\_1\_1750\_@runat\_05052003\_173955\_334 -ICPU=3  
-NO\_XLIFY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_cm.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pi.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rod.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgtgo\_hum.\*  
40: em\_hgtgo\_mus.\*  
41: em\_hgtgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	532	100.0	337	6	AX321911 Sequence
2	527	99.1	5981	6	AX321910 Sequence
3	513	96.4	2239	6	AX321909 Sequence
c 4	322	60.5	161280	9	AC093903 Homo sapi
5	221	41.5	575	6	AX341060 Sequence
c 6	221	41.5	596	6	AX351341 Sequence
7	221	41.5	1683	9	AB042201 Homo sapi
8	221	41.5	1861	9	AB026891 Homo sapi
9	221	41.5	1874	9	AF200708 Homo sapi
10	221	41.5	2000	9	AB040875 Homo sapi
11	221	41.5	2155	9	BC012087 Homo sapi
12	221	41.5	2482	9	AF252872 Homo sapi
13	210.5	39.6	3144	9	HS277882 Homo sapi
14	144.5	27.2	2216	10	AB022345 Mus muscu
15	144.5	27.2	2316	10	AB037650 Mus muscu
16	131	24.6	96640	2	AC106184 Rattus no
17	131	24.6	183949	2	AC118181 Rattus no
18	131	24.6	202985	2	AC127139 Rattus no
c 19	114	21.4	189747	2	AC087791 Homo sapi
20	112.5	21.1	190815	9	AC007066 Homo sapi
21	112.5	21.1	190824	2	AL731645 Homo sapi
22	112	21.1	155025	9	AC034246 Homo sapi
23	112	21.1	155365	9	AC107970 Homo sapi
24	112	21.1	168720	2	AC078781 Homo sapi
25	112	21.1	194334	2	AC046140 Homo sapi
c 26	112	21.1	197110	9	AC104306 Homo sapi
c 27	111.5	21.0	182509	9	AC112211 Homo sapi
c 28	111	20.9	107795	9	AL662873 Human DNA
29	111	20.9	149850	9	AP001622 Homo sapi
30	111	20.9	340000	9	AP001746 Homo sapi
31	110.5	20.8	110000	2	AC080008 Homo sapi
32	110.5	20.8	215077	9	AC112504 Homo sapi
c 33	110	20.7	118819	9	AC011491 Homo sapi
34	109.5	20.6	39958	9	AC011542 Homo sapi
35	109.5	20.6	41716	9	AC011522 Homo sapi
c 36	109	20.5	41572	2	AC006132 Homo sapi
c 37	109	20.5	77322	2	AC021334 Homo sapi
c 38	109	20.5	127361	9	AC007283 Homo sapi
c 39	109	20.5	145414	9	HS392M18 Human DNA
c 40	109	20.5	179216	2	AC074016 Homo sapi
c 41	109	20.5	184063	9	AC023481 Homo sapi
42	108.5	20.4	124086	9	AC008882 Homo sapi
c 43	108.5	20.4	127171	2	AC008777 Homo sapi
c 44	108.5	20.4	145264	9	AC016601 Homo sapi
c 45	108.5	20.4	150296	9	U91326 Homo sapien

ALIGNMENTS

RESULT 1

```
AX321911
LOCUS AX321911 337 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 442 from Patent WO0172295.
ACCESSION AX321911
VERSION AX321911.1 GI:17906521
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indrias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 442 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..337
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 103 a 60 c 93 g 81 t
ORIGIN
Alignment Scores:
Pred. No.: 1.72e-55 Length: 337
Score: 532.00 Matches: 97
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-854-133-586 (1-97) x AX321911 (1-337)
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
Db 5 GAGTTGAAGTGAGCAGAGATCATGCCAGCTGGGTGACGTGACATCTGCTCAACAA 64
QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40
Db 65 GAATAGGAAAAAAGAAAGAAAGAAAGAGAGAGAGAAATCCAGGCCAAATTGTGTC 124
QY 41 IleAspPheLlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGln 60
Db 125 ATAGATTATATCATATCTGGATTTTGGATTTCTTGTCTCATCATCTGGATTCAG 184
QY 61 GluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGly 80
Db 185 GAAGCTGTGTGTCACCATCTCCAAAGAGAGTTACCTGCAGGAAATGTTAACGGGA 244
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97
Db 245 GGCTGCTTCCCTGGGCAACAGAGGCCACCTGGGAGGAGAAAGTGCGAGC 295
RESULT 2
AX321910
LOCUS AX321910 5981 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 441 from Patent WO0172295.
ACCESSION AX321910
VERSION AX321910.1 GI:17906518
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indrias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 441 04-OCT-2001;
```

```
FEATURES
source Location/Qualifiers
1..5981
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1858 a 1029 c 1098 g 1996 t
ORIGIN
Alignment Scores:
Pred. No.: 1.82e-53 Length: 5981
Score: 527.00 Matches: 96
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.06% Indels: 0
DB: 6 Gaps: 0
US-09-854-133-586 (1-97) x AX321910 (1-5981)
QY 2 ValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGlu 21
Db 3 GTTGAAGTGAGCAGAGATCATGCCAGCTGGGTGACAGTGACACTCTGCTCAACAGAA 62
QY 22 LeuArgLysLysGluArgLysLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 41
Db 63 TTAAGGAAAAAAGAAAGAAAGAAAGAGAGAGAGAAATCCAGGCCAAATTGTGCGATA 122
QY 42 AspPheLleLlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGlnGlu 61
Db 123 GATTTTATCATATCTGGATTTTGGATTTTGTGTTTCTCATCATCTGGATTCAGAA 182
QY 62 SerLeuLeuCysProProSerProLysGluValThrCysArgGluMetLeuThrGlyGly 81
Db 183 AGCCTGTGTGTCACCATCTCCAAAGAGAGGTACCTGCAGGAAATGTTAACGGGAGGC 242
QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97
Db 243 TGCTTCCCTGGGCAACAGAGGCCACCTGGGAGGAGAAAGTGCGAGC 290
RESULT 3
AX321909
LOCUS AX321909 2239 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 440 from Patent WO0172295.
ACCESSION AX321909
VERSION AX321909.1 GI:17906515
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indrias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 440 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..2239
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 619 a 444 c 493 g 683 t
ORIGIN
Alignment Scores:
Pred. No.: 2.97e-52 Length: 2239
Score: 513.00 Matches: 97
Percent Similarity: 84.35% Conservative: 0
Best Local Similarity: 84.35% Mismatches: 0
Query Match: 96.43% Indels: 18
DB: 6 Gaps: 1
US-09-854-133-586 (1-97) x AX321909 (1-2239)
```



QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 Db 2 GAGGTTGAAGTGAAGCAGAGATCATGCCAGCTGGGTGACAGTCTCTCTCAACA 61  
 QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgGluArgLysPheGlnAlaAsnCysGly 40  
 Db 62 GAATTAAGGAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATTCACAGGCCAATTTGTGGC 121  
 QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGln 60  
 Db 122 ATAGATTATCATATCTGGAATTTTGGATTCTTTTGTCTCTCATCTGGATTTCAG 181  
 QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 Db 182 GAAAGGCTGTGTCTCCACCATCTCCAAAGGAGGTACCTGCAGGGAATTTTACGGGA 241  
 QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg----- 93  
 Db 242 GGCTGCTTCCCTGGGCAACAGAGACCCTGGCAGGAGCCCTTTTCAGGAGAGACG 301  
 QY 94 -----ArgLysCysSer 97  
 Db 302 CCTTTTCAGGAGAGACGCCCTTTTCAGGAGAGAGAAAGTGCAGC 346  
 RESULT 4  
 AC093903/c  
 LOCUS Homo sapiens BAC clone RP11-733C7 from 4, complete sequence.  
 DEFINITION AC093903 AC055827  
 ACCESSION AC093903.3 GI:15920156  
 VERSION HTG.  
 KEYWORDS Homo sapiens.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Sulston, J.E. and Waterston, R.  
 1 (bases 1 to 161280)  
 Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 161280)  
 Radonenko, M. and Kozlowski, A.  
 The sequence of Homo sapiens BAC clone RP11-733C7  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 161280)  
 Waterston, R.H.  
 Direct Submission  
 TITLE Submitted (10-SEP-2001) Genome Sequencing Center, Washington  
 JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 161280)  
 Waterston, R.H.  
 Direct Submission  
 TITLE Submitted (04-OCT-2001) Genome Sequencing Center, Washington  
 JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 161280)  
 Waterston, R.H.  
 Direct Submission  
 TITLE Submitted (01-MAR-2002) Department of Genetics, Washington  
 JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Oct 4, 2001 this sequence version replaced gi:15625016.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0733C07  
 Drafting Center: WIBR  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-310A13. Actual start of this clone is at base position 1 of RP11-733C7; actual end is at base position 161280 of RP11-733C7.

Data from AC009792 was used to finish the clone, AC055827.

The sequence of AC055827 has been incorporated into AC093903.

FEATURES	source	Location/Qualifiers
repeat_region	1. .161280	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="4" /map="4" /clone="RP11-733C7" /clone_lib="RPC1-11" /rpt_family="L1" 306. .406
repeat_region	306. .406	/rpt_family="(GGAA)n" 471. .525
repeat_region	471. .525	/rpt_family="AT-rich" 530. .782
repeat_region	530. .782	/rpt_family="L1" 685. .712
repeat_region	685. .712	/rpt_family="(TA)n" 1383. .1472
repeat_region	1383. .1472	/rpt_family="GA-rich" 1697. .1836
repeat_region	1697. .1836	/rpt_family="L1" 1837. .1956
repeat_region	1837. .1956	/rpt_family="L1" 3005. .3221
repeat_region	3005. .3221	/rpt_family="MIR" 4013. .4321
repeat_region	4013. .4321	/rpt_family="Alu" 4295. .4325
repeat_region	4295. .4325	/rpt_family="(A)n" 5625. .5960
repeat_region	5625. .5960	/rpt_family="L1" 5938. .5982
repeat_region	5938. .5982	/rpt_family="A-rich" 6062. .6195



Qy 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
 |||||  
 Db 256 -----GAAAGCGTGTGTGTCACCATCTCCAAAGAGAGGTTACCTGCAGGAAATG 306  
 |||||  
 Qy 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgGlyCysSer 97  
 |||||  
 Db 307 TTAACGGGAGGCTGCTTCCCTGGGCAACAGAGCCACTGGGCAGGAGAAAGTGCAGC 366  
 |||||  
 RESULT 6  
 AX351341/c AX351341 596 bp DNA linear PAT 06-FEB-2002  
 LOCUS  
 DEFINITION Sequence 88 from Patent WO0196390.  
 ACCESSION AX351341  
 VERSION AX351341.1 GI:18616688  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Jiang,Y., Hepler,W.T., Clapper,J.D., Wang,A. and Secrist,H.  
 Compositions and methods for the therapy and diagnosis of colon  
 cancer  
 JOURNAL Patent: WO 0196390-A 88 20-DEC-2001;  
 CORIXA CORPORATION (US)  
 FEATURES Location/Qualifiers  
 source  
 1..596  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 160 a 151 c 126 g 158 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,67e-17 Length: 596  
 Score: 221.00 Matches: 47  
 Percent Similarity: 70.00% Conservative: 9  
 Best Local Similarity: 58.75% Mismatches: 18  
 Query Match: 41.54% Indels: 6  
 DB: 2  
 Gaps: 2  
 US-09-854-133-586 (1-97) x AX351341 (1-596)  
 Qy 19 GlnThrGluLeuArgLysLysGluArgLys---LysLysArgGluValThrCysArgGluMet 37  
 |||||  
 Db 481 CAGTCGAAGCAGAGGAGACATCGATCAGTAACACCAAGACACCAAGTTGAAGT 422  
 |||||  
 Qy 38 AsnCysGlyLeuAspPheIleLeuPheTrpIlePheTrpIleLeuLeuPheSerHisHis 57  
 |||||  
 Db 421 TTTGTTTCTTCCCTGCTGTTTATTTTCCCGCGTGTGTCCTTACTATGTCA----- 368  
 |||||  
 Qy 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
 |||||  
 Db 367 -----GAAAGCGTGTGTGTCACCATCTCCAAAGAGGTTACCTGCAGGAAATG 317  
 |||||  
 Qy 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgGlyCysSer 97  
 |||||  
 Db 316 TTAACGGGAGGCTGCTTCCCTGGGCAACAGAGCCACTGGGCAGGAGAAAGTGCAGC 257  
 |||||  
 RESULT 7  
 AB042201 1683 bp DNA linear PRI 10-FEB-2001  
 LOCUS  
 DEFINITION Homo sapiens xCT gene for cystine/glutamate transporter, exon 1,  
 partial cds.  
 ACCESSION AB042201  
 VERSION AB042201.1 GI:7670267  
 KEYWORDS cystine/glutamate transporter.  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (sites)  
 Sato,H., Tamba,M., Kuriyama-Matsumura,K., Okuno,S. and Bannai,S.  
 Molecular cloning and expression of human xCT, the light chain of

amino acid transport system xc-  
 Antioxidants and Redox Signaling 2, 665-671 (2000)  
 JOURNAL  
 REFERENCE 2 (bases 1 to 1683)  
 AUTHORS Sato,H. and Bannai,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-APR-2000) Hideyo Sato, University of Tsukuba,  
 Institute of Basic Medical Sciences; Tennodai 1-1-1, Tsukuba,  
 Ibaraki 305-8575, Japan (E-mail:hideyo-semd.tsukuba.ac.jp,  
 Tel:81-298-53-3282, Fax:81-298-53-3039)  
 FEATURES Location/Qualifiers  
 source  
 1..1683  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 gene 1001..1683  
 /gene="xCT"  
 exon 1001..1555  
 /gene="xCT"  
 /product="cystine/glutamate transporter"  
 CDS 1279..>1555  
 /number=1  
 /gene="xCT"  
 /codon\_start=1  
 /product="cystine/glutamate transporter"  
 /protein\_id="BAA94999.1"  
 /db\_xref="GI:7670268"  
 /translation="MVRPVPVSTISKGGYLOGNVNRPLPSLGNKEPPGQKVKRKV  
 TLRGVSTIIIGIIGAGIFISPKGVLTQNTSGVMSLTITWVCGVLSLF"  
 1556..>1683  
 /gene="xCT"  
 /number=1  
 intron  
 BASE COUNT 495 a 287 c 370 g 531 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 5,41e-17 Length: 1683  
 Score: 221.00 Matches: 47  
 Percent Similarity: 70.00% Conservative: 9  
 Best Local Similarity: 58.75% Mismatches: 18  
 Query Match: 41.54% Indels: 6  
 DB: 2  
 Gaps: 2  
 US-09-854-133-586 (1-97) x AB042201 (1-1683)  
 Qy 19 GlnThrGluLeuArgLysLysGluArgLys---LysLysArgGluValThrCysArgGluMet 37  
 |||||  
 Db 1172 CAGTCGAAGCAGAGGAGACATCGATCAGTAACACCAAGACACCAAGTTGAAGT 1231  
 |||||  
 Qy 38 AsnCysGlyLeuAspPheIleLeuPheTrpIlePheTrpIleLeuLeuPheSerHisHis 57  
 |||||  
 Db 1232 TTTGTTTCTTCCCTGCTGTTTATTTTCCCGCGTGTGTCCTTACTATGTCA----- 1285  
 |||||  
 Qy 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77  
 |||||  
 Db 1286 -----GAAAGCGTGTGTGTCACCATCTCCAAAGAGGTTACCTGCAGGAAATG 1336  
 |||||  
 Qy 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgGlyCysSer 97  
 |||||  
 Db 1337 TTAACGGGAGGCTGCTTCCCTGGGCAACAGAGCCACTGGGCAGGAGAAAGTGCAGC 1396  
 |||||  
 RESULT 8  
 AB026891 1861 bp mRNA linear PRI 10-FEB-2001  
 LOCUS  
 DEFINITION Homo sapiens mRNA for cystine/glutamate transporter, complete cds.  
 ACCESSION AB026891  
 VERSION AB026891.1 GI:5668544  
 KEYWORDS cystine/glutamate transporter.  
 SOURCE Homo sapiens fibroblast cDNA to mRNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (sites)  
 Sato,H., Tamba,M., Kuriyama-Matsumura,K., Okuno,S. and Bannai,S.  
 Molecular cloning and expression of human xCT, the light chain of



**AUTHORS** Kim, J. Y., Chairoungdua, A., Cha, S. H., Segawa, H., Matsuo, H., Kim, D. K., Endou, H. and Kanai, Y.

**TITLE** Human cystine/glutamate exchanger: cDNA cloning and upregulation by oxidative stress in glioma cells

**JOURNAL** Unpublished

**REFERENCE** 2 (bases 1 to 2000)

**AUTHORS** Kanai, Y.

**TITLE** Direct Submission

**JOURNAL** Submitted (01-APR-2000) Yoshikatsu Kanai, Kyorin University School of Medicine, Department of Pharmacology and Toxicology; 6-20-2 Shinkawa, Mitaka, Tokyo 181-8611, Japan  
(E-mail: ykanai@kyorin-u.ac.jp, Tel: +81-422-47-5511(ex.3453), Fax: +81-422-79-1321)

FEATURES	Location/Qualifiers
Source	1..2000 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="hxCT" /clone_lib="brain cDNA library" /dev_stage="adult"
gene	1..2000 /gene="hxCT"
CDS	136..1620 /gene="hxCT" /note="Na+-independent amino acid transporter transmembrane protein" /codon_start=1 /product-"cystine/glutamate exchanger" /protein_id="BAB40574.1" /db_xref="GI:13516846" /translation="MVRRKDVSTISKGYLQGVNVRSLPSLNKPPEPGOEKVQLKKRV TLLRGVSIITIGIIGAGIFISPKGVLQNTSGVMGLTIWTCGVLISFGALSYAEILGT TIKKSGHYTYILEVGPPLPAFVRVWVELLIIRPATAVISLAFRXYILEPFPIQEI PELAIKTAGIVTVVNLNSMSWSARIQIFLTFCKLTAIDLIIVPGVMQLKGOT QNKDAFSDGSIIITRLPLAFYGYAGWFYNVEETVENPERTIIPALICISAKI VIYGVLINVAFTIENABELLSNAVATFGERLGNFSLAVPIFVALSCFSGMNG VFASRLFFVASREGHLPIILSMIHVRKTPLPNAVILHPLTMILFSGDLDSLALLF SPARWLFIGLAVGLIYLRYKCPDMHRFPKVPLEIALFSPTCLEWVALSLYSDPST GIGFVITLTGVPAYLYFIWDKAPRFMRMSGFLALMPAQADM"
BASE COUNT	505 a 418 c 435 g 642 t
ORIGIN	
Alignment Scores:	
Pred. No.:	6.58e-17 Length: 2000
Score:	221.00 Matches: 47
Percent Similarity:	70.00% Conservative: 9
Best Local Similarity:	58.75% Mismatches: 18
Query Match:	41.54% Indels: 6
DB:	9 Gaps: 2

US-09-854-133-586 (1-97) x AB040875 (1-2000)					
QY	19	GlnThrGluLeuArgLysLysGluArgLys---	LysLysArgGluArgLysPheGlnAla	37	
		:		:	:
Db	29	CAGTCTGAAGACGAGGAAACATCGATCAGTAACCAACAGACACACAAGTTGAAGT	88		
		:		:	:
QY	38	AsnCysGlyIleAspPheIleIlePheTrpIleLeuLeuPheSerHisHis	57		
		:  :::  ::	:	:	:
Db	89	TTTGTTTTCTTCCCTCTGTATTATTTTCCCCCGTGTCCCTACTATGTGTC	142		
		:		:	:
QY	58	TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet	77		
		:		:	:
Db	143	-----GAAAGCCTGGTTGTCTCCACCATCTCCAAGGAGGTTTACCTCAGGAGGAATG	193		
		:		:	:
QY	78	LeuThrGlyClyCysLeuProTprAlaThrArgSerHisLeuGlyArgArgLysCysSer	97		
		:		:	:
Db	194	TTAACGGGAGGCTGCCTTCCTCGGCCAACAAGGAGCACCTGGCAGGAGAAAATGCGAG	253		
		:		:	:
<b>RESULT 11</b>					
BC012087					
LOCUS					
DEFINITION					
BC012087 Homo sapiens, similar to solute carrier family 7, (cationic amino acid transporter, y+ system) member 11, clone MGC:20026					

IMAGE:4562394, mRNA, complete cds.	
BC012087	
VERSION	GI:15082351
KEYWORDS	MGC.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 2155)
JOURNAL	Direct Submission
	Straussberg, R.
	Submitted (02-AUG-2001) National Institutes of Health, Mammalian
	Gene Collection (MGC), Cancer Genomics Office, National Cancer
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
	USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk
	Email: <a href="mailto:cgabps-r@mail.nih.gov">cgabps-r@mail.nih.gov</a>
	Tissue Procurement: DCTD/DTP
	cDNA Library Preparation: Rubin Laboratory
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: Genome Sequence Centre,
	BC Cancer Agency, Vancouver, BC, Canada
	<a href="mailto:info@bcgsc.bc.ca">info@bcgsc.bc.ca</a>
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
	Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
	Leticia Hsiao, Martin Krzywinski, Reta Kutscher, Oliver Lee, Soo
	Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
	Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacqueline
	Schein, Duane Smillaus, Michael Smith, Lorraine Spence, Jeff Stott,
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
	George Yang, Scott Zuyderduyn, Marco Marra.
	Clone distribution: MGC clone distribution information can be found
	through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>
	Series: IRAL Plate: 29 Row: g Column: 1
	This clone was selected for full length sequencing because it
	passed the following selection criteria: matched mRNAqi: 5668544.

```

FEATURES
source
Location/Qualifiers
1. .2155
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:20026 IMAGE:4562994"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
161. .1666
CDS
/codon_start=1
/product="Similar to solute carrier family 7, (cationic
amino acid transporter, y+ system) member 11"
/protein_id="AAH12087.1"
/db_xref="GI:15082352"
/translation="MYRKPVSTISRGYLGQNVNRLPSLGNKEPPGQEKVLKRV
TLIRGSLIIGITIGAGITISPKGVLQNTSGVMSLTIWTCGVLSLFGALSALVEL
TIKSGGHTYILEVGPPLPAFYRVVVELLIIRPAATAVISLAFGRVILPFFIOCEI
PELAKLIITAVGTIVYVNLNSKYSWSARIQIELTECKLTAILIILVPGVMOLIKGQT
QNFKAFSGDSITRLPGLTYGTAYAGFYLNFTVEENPKTIPALICISWAI
TVIGVYLNTWATYITINAEELLNSNAVITFSERLIGNFLSLAVPIFVALSCFSMGNG
VFVASFVLFVARSREGLPLTSLIMHYKHTPLPVAIVLPLTMIFLFGDLSLALFL
SPARWFLIGLAVAGLYLYIAKCPDMRHPKVPILFIPALFSFTCLFWALSILYSDPFT
GIGFVITLGVPAYLFIYIWDKPKRFIMSEKITITLQIILEVVPEDKL"
BASE COUNT 599 a 443 c 452 g 561 t
ORIGIN
Alignment Scores:
Pred. No.: 7,16e-17
Score: 221.00
Percent Similarity: 70.00%
Best Local Similarity: 58.75%
Query Match: 41.54%
DB: 9

```

Alignment Scores:	
Pred. No.:	7,16e-17
Score:	221.00
Percent Similarity:	70.00%
Best Local Similarity:	58.75%
Query Match:	41.54%
DB:	9
Length:	2155
Matches:	47
Conservative:	9
Mismatches:	18
Indels:	6
Gaps:	2

QY 19 GlnThrGluLeuArgLysLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37

Db	125	CAGTCTGAACACGAGGAAGACATCGATCATGACATACCAACAGAGACACCAAGTTGAAGT	184
Qy	38	AsnCysGlyIleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHis	57
Db	185	TTTGTGTTCTTCCTCTGTGTTTATTTTCCCGGTGTGTCCTACTATGGTCA	238
Qy	58	TrpIleGlnGluSerLeuLeuCysProSerProLysGluValThrCysArgGluMet	77
Db	239	-----GAAAGCCCTGTGTGTCCACCATCTCCAAGGAGGTTACTCTCAGGAAATG	289
Qy	78	LeuThrGlyGlyCysLeuProTirPAlaThrArgSerHisLeuGlyArgLysCysSer	97
Db	290	TTACGGGAGGCTCCCTTCCTGGGCAACAAGGAGCACCTGGCAGAGAAAGTGCGAGC	349
RESULT	13		
LOCUS	HSA277882	3144 bp	linear
DEFINITION	Homo sapiens mRNA for cystine/glutamate transporter (XCT gene).		
ACCESSION	AJ277882		
VERSION	AJ277882.1	GI:18073361	
KEYWORDS	cystine/glutamate transporter; XCT gene.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 Borsani,G., Manzoni,M., Palacin,M., Pineda,M. and Gasol,E.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 3144)		
AUTHORS	Bassi,M.T.		
TITLE	Direct Submition		
JOURNAL	Submitted (11-MAY-2000) Bassi M.T., Telethon Institute of Genetics and Medicine, Via Olgettina 58, 20132 Milan, ITALY		
FEATURES	Location/Qualifiers		
source	1..3144		
gene	/organism="Homo sapiens"		
CDS	/db_xref="taxon:9606"		
	/chromosome="4"		
	/map="4q28-q32"		
	/cell_line="NT2, undifferentiated teratocarcinoma cell line"		
	1..3128		
	/gene="XCT"		
	1..1506		
	/gene="XCT"		
	/codon_start=1		
	/product="cystine/glutamate transporter"		
	/protein_id="CAC81305.1"		
	/db_xref="GI:18073362"		
	/translation="MYRKVIVSTISKGYLQGNVGRPLSLGNKPPGQEKVOLKRRKVTLLRGSYIIITGIAGIPISIPKGVQLNTGSVMSLTIWTGVLFLGALSYAEIQT		
	TIKSGHGTYILEVGFYPAFVRVMSVQIIIRPAATAVISLAFGRYILEPFFIOCEI		
	PELAKLITAGVTVMVLNSMSVRSQILFTFKLTAILIIVPGVMOLIKGOT		
	QNEKFAFGSDSITRLPLAFYGVYAVAGWFLNVEVENPEKTIPLAICISMAI		
	VIQGYLVNVAFTTINAEILLISNAVATVSERLLGNFSLAVIPVALSCFGSMNG		
	VFASVRLTVASREGHLPISIMHVKRTPIPLAVILHPLTMIMLFSGDLSLNLFI		
	SFARWFLIGLAVAGLRYLKCPDMHRPEKVPFLTFPALFSFTCLFMVALSLYSDPFS		
	TGCFVITLGVGPAYYFLIWDKKPRFRIMSEKITRTLQILLEVVPEDKL"		
repeat_region	2857..3127		
polyA_site	3128		
	/rpt_family="Sine/Alu"		
BASE COUNT	912 a	602 c	649 g 981 t
ORIGIN			
Alignment Scores:			
Pred. No.:	2,06e-15	Length:	3144
Score:	210.50	Matches:	38
Percent Similarity:	95.00%	Conservative:	0
Best Local Similarity:	95.00%	Mismatches:	1
Query Match:	39.57%	Indels:	1
DB:	9	Gaps:	1

US-09-854-133-586 (1-97) x HSA277882 (1-3144)

QY 58 TrpIleGlnGluSerLeuLeuCysProSerProLysGluValThrCysArgGluMet 77  
 Db 2 TGG---TCAGAAAGCGCTGTTGTCACCACTCCAAAGAGGTTACCTGCAGGGAATG 58  
 QY 78 LeuThrGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgGlyCysSer 97  
 Db 59 TTAACGGGAGGCTGCTTCCCTGGGCAACAGGAGCCACCTGGGCGAGGAAATGCAGC 118

RESULT 14  
 AB022345 2216 bp mRNA linear ROD 27-APR-1999  
 LOCUS Mus musculus xct mRNA for cystine/glutamate transporter, complete cds.  
 DEFINITION  
 AB022345 2216 bp mRNA for cystine/glutamate transporter, complete cds.  
 ACCESSION  
 AB022345 1 GI:4689080  
 VERSION  
 xct; cystine/glutamate transporter.  
 KEYWORDS  
 Mus musculus cDNA to mRNA.  
 SOURCE  
 Mus musculus  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 (sites)  
 Sato,H., Tamba,M., Ishii,T. and Bannai,S.  
 Cloning and expression of a plasma membrane cystine/glutamate exchange transporter composed of two distinct proteins  
 J. Biol. Chem. 274 (17), 11455-11458 (1999)  
 J9223452

REFERENCE  
 2 (bases 1 to 2216)  
 Sato,H. and Bannai,S.  
 Direct Submission  
 Submitted (13-JAN-1999) Hideyo Sato, University of Tsukuba,  
 Institute of Basic Medical Sciences; Tennodai 1-1-1, Tsukuba,  
 Ibaraki 305-8575, Japan (E-mail:hideyo.smd.tsukuba.ac.jp,  
 Tel:81-298-53-3282, Fax:81-298-53-3039)

FEATURES  
 source  
 1. .2216  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 1. .2216  
 /gene="xct"  
 320. .1828  
 /gene="xct"  
 /function="anionic amino acid transport"  
 /evidence=experimental  
 /product="cystine/glutamate transporter"  
 /protein\_id="BAA77220.1"  
 /db\_xref="GI:4689081"  
 /translation="MVRKPVVATISKGYLQGNMGRPLSMGDDPPGQEKVVLKKKI  
 TLRGVSIITIGVIGIFSPKILQNTGVSGLVSWACGVLISLFGALSVAELGT  
 SIKSGGHVYILEVFGPLAFVWVVELLVRPGATAVISLAFGRYILEPFIQCEI  
 PELAKLVAVGTVMVNLSTSVSWARQIFLTFCCKLTAIIIIIVPGVQLIKQPT  
 HFKDASGSDTSLMGLPLAFYAGVWFLNFITEEDVNDPEKTIPLAICISMAI  
 ITGVYLVNVAFTTISAEELQSSAVATFSERLKGLSLAVFIALSCFGSMNG  
 VFASVFLFYASREGLHPEILSMTHVHKHPLPAVILHPLTMVLMFSGDLYSLNLF  
 SFARWFLMGLAVAGLIYLRKPDHMRPKVPLFIPALFSFTCLFMVVLVSLSDPFT  
 GVGFLLTGTGPAYLFIIVWKKPKFWRSLDRITRTIQLIILEVVPDSKEL"  
 2216  
 /gene="xct"  
 /notes="34 a nucleotides"  
 542 a 526 c 509 g 639 t

polyA\_site  
 BASE COUNT 542 a 526 c 509 g 639 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.39e-07 Length: 2216  
 Score: 144.50 Matches: 38  
 Percent Similarity: 51.06% Conservative: 10  
 Best Local Similarity: 40.43% Mismatches: 33  
 Query Match: 27.16% Indels: 13  
 DB: 10 Gaps: 3

US-09-854-133-586 (1-97) x AB022345 (1-2216)  
 QY 11 LeuGlyAspSerGluThrLeuSerGlnThrGluLeuArgLysLysGluArgLysLysLys 30  
 Db 162 GTGGTGATACGAAGGGGAAGTCACGACCAAGACAGTACATCAGTCACCTCTTAGAGAAC 221  
 QY 31 ArgGluArgLysPheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 50  
 Db 222 AAGTAAAGGGTGTGTTTGTGTTTGT-----TTTATTTTCTCTGTTTGTGTTT 272  
 QY 51 IleLeuLeuPheSerHisHis-----TrpIleGlnGlu 61  
 Db 273 CCCCTCTGTTTCTTTTTCATCCCTCCTCTGCTGTGTGACACTGCCATGG---TCAGAA 329  
 QY 62 SerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGlyGly 81  
 Db 330 AGCAGTTGTGGCCACCATCTCCAAAGAGGTTTACCTCAGGGCAATATGACGGGAGGC 389  
 QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgLys 95  
 Db 390 TGCCCTCCATGGGGGACCAAGAGCCACCTGGGAGGAGG 431

RESULT 15  
 AB03765S01  
 LOCUS  
 DEFINITION  
 AB037650  
 ACCESSION  
 AB037650.1 GI:6863057  
 VERSION  
 xct.  
 KEYWORDS  
 1 of 12  
 SEGMENT  
 SOURCE  
 Mus musculus (strain:129SVJ) female liver DNA.  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1  
 Sato,H., Kuriyama-Matsumura,K., Hashimoto,T., Sasaki,H., Wang,H.,  
 Ishii,T., Mann,G.E. and Bannai,S.  
 Effect of oxygen on induction of the cystine transporter by  
 bacterial lipopolysaccharide in mouse peritoneal macrophages  
 J. Biol. Chem. 276 (13), 10407-10412 (2001)  
 21167854  
 2 (bases 1 to 2316)  
 Sato,H., Sasaki,H. and Bannai,S.  
 Direct Submission  
 Submitted (26-JAN-2000) Hideyo Sato, University of Tsukuba,  
 Institute of Basic Medical Sciences; Tennodai 1-1-1, Tsukuba,  
 Ibaraki 305-8575, Japan (E-mail:hideyo.smd.tsukuba.ac.jp,  
 Tel:81-298-53-3282, Fax:81-298-53-3039)

FEATURES  
 source  
 1. .2316  
 /organism="Mus musculus"  
 /strain="129SVJ"  
 /db\_xref="taxon:10090"  
 /sex="female"  
 /tissue\_type="liver"  
 1510. .2105  
 /gene="xct"  
 /number=1  
 627 a 462 c 509 g 718 t

BASE COUNT 627 a 462 c 509 g 718 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.46e-07 Length: 2316  
 Score: 144.50 Matches: 38  
 Percent Similarity: 51.06% Conservative: 10  
 Best Local Similarity: 40.43% Mismatches: 33  
 Query Match: 27.16% Indels: 13  
 DB: 10 Gaps: 3

US-09-854-133-586 (1-97) x AB03765S01 (1-2316)

QY 11 LeuGlyAspSerGluThrLeuSerGlnThrGluLeuArgLysLysGluArgLysLysLys 30  
 ::::::::::::::::::::

Db 1671 GTCGGTGATAGCAAGGGGAAGTACAGCCGAACAGTATGATCAGTCACTTCTTAGAGAAAC 1730  
QY 31 ArgGluArgLysPheGlnAlaAsnCysGlyIleAaspPheIleIlePheTrpIlePheTrp 50  
Db 1731 AAGTTAAAGGGTTTGTGTTTGT-----TTTATTTGCTGTTGTTTGTGTTT 1781  
QY 51 IleLeuLeuPheSerHisHis-----TrpIleGlnGlu 61  
Db 1782 CCCCTCTGTGTTTCTTTTTCATCCCTCTGCTGTGACACTGCCATGG---TCAGAA 1838  
QY 62 SerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGlyGly 81  
Db 1839 ACCAGTTGTGCCACCACTCTCCAAGGAGGTACCTGCAGGCAATATGACGGGAGGC 1898  
QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLys 95  
Db 1899 TGCCCTCCATGGGGACCAAGAGCCACCTTGGGCAGGAGAAGG 1940

Search completed: May 11, 2003, 14:59:28  
Job time : 2223.24 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 12:41:29 ; Search time 210.31 Seconds  
(without alignments)  
1038.677 Million cell updates/sec

Title: US-09-854-133-586

Perfect score: 532

Sequence: 1 EVEVSRDHASLGDSFSLTQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_p2n.model -DRV=xlh  
-Q/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_173955\_327/app\_query.fasta\_1.462  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133 @CEN\_1\_1\_275 @runat\_05052003\_173955\_327 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	532	100.0	337	23	AAD23462	Human lung tumour-
2	527	99.1	5981	23	AAD23461	Human lung tumour-
3	513	96.4	2239	23	AAD23460	Human lung tumour-
4	221	41.5	575	24	ABL37718	Human colon cancer
5	221	41.5	596	24	ABK27651	Human gene express
6	221	41.5	800	20	AAZ16609	Human cysteine/Glu
7	211.5	39.8	1958	22	ABA09201	Human EST-derived
8	211.5	39.8	1958	22	AAK98573	Human immune/haema
9	206.5	38.8	123	22	AAK68639	Human immune/haema
10	104.5	19.6	18408	22	AAAL04753	Human reproductive
11	104.5	19.6	18408	22	AAAL04754	Human reproductive
12	104.5	19.6	18408	23	ABL97656	Human testicular a
13	104.5	19.6	18408	23	ABL97657	Human testicular a
14	102	19.2	971	21	AAK58883	Human tumour suppr
15	100	18.8	3424	22	AAAL06199	Human reproductive
16	100	18.8	3424	23	ABL98764	Human testicular a
17	99	18.6	109906	24	ABK94411	DNA encoding endot
18	98.5	18.5	4013	22	AAK68715	Human immune/haema
19	98.5	18.5	23164	22	AAK79678	Human immune/haema
20	97.5	18.3	154465	24	AAD28763	Human AKAP allelic
21	97.5	18.3	158245	24	AAD28762	Human AKAP allelic
22	97.5	18.3	161425	22	AAH02340	Human AKAP10 gene
23	97.5	18.3	162025	22	AAH02339	Human AKAP10 gene
24	97.5	18.3	162025	24	AAD28758	Human AKAP allelic
25	97.5	18.3	162025	24	AAD28759	Human AKAP allelic
26	97	18.2	178896	24	ABO88146	Human osteoblast d
27	96.5	18.1	9358	24	ABK84319	Human cDNA differe
28	96.5	18.1	9358	24	ABL61905	Colon adenocarcino
29	96	18.0	32249	22	ABAI7155	Human nervous syst
30	96	18.0	107820	22	AAD16230	Human ATP-binding
31	95.5	18.0	6250	22	AAK79549	Human immune/haema
32	95	17.9	752	22	AAK36489	Human cardiovascular
33	95	17.9	2138	22	AAAL06186	Human reproductive
34	95	17.9	2138	23	ABL98751	Human testicular a
35	94.5	17.8	160552	22	AAO2697	Human glycosyl sul
36	94	17.7	2824	21	AAK59843	Human secreted pro
37	94	17.7	38771	22	AAK81036	Human immune/haema
38	93.5	17.6	18132	22	AAK67033	Human immune/haema
39	93.5	17.6	18132	22	AAK79622	Human immune/haema
40	93.5	17.6	47319	22	AAK64813	Human immune/haema
41	93.5	17.6	47319	22	AAK72230	Human immune/haema
42	93.5	17.6	236303	22	AAH11614	Human genomic DNA
43	93	17.5	2591	22	AAH18566	Human cDNA sequenc
44	93	17.5	28690	22	AAF57718	Human IL4Ralpha ge
45	93	17.5	31813	22	ABA08137	Human ovarian and

## ALIGNMENTS

RESULT 1  
AAD23462  
ID AAD23462 standard; cDNA; 337 BP.  
XX AAD23462;  
AC AAD23462;  
XX  
XX 26-FEB-2002 (first entry)  
DT Human lung tumour-specific 20E10 5' cDNA.  
DE  
XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
KW antisense-therapy; vaccine; immune response; lung cancer; 20E10; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200172295-A2.  
XX  
XX 04-OCT-2001.  
PD

XX 28-MAR-2001; 2001WO-US09991.  
 XX 29-MAR-2000; 2000US-0538037.  
 PR 05-JUN-2000; 2000US-0588937.  
 PR 18-AUG-2000; 2000US-0640878.  
 PR 22-SEP-2000; 2000US-234517P.  
 PR 01-NOV-2000; 2000US-0704512.  
 PR 14-DEC-2000; 2000US-0738973.  
 XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Mohamath R, Secrlist H, Benson DR, Indirias CY;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX WPI; 2001-639201/73.

XX New human lung-specific polynucleotides and polypeptides for the  
 PT diagnosis and treatment of disease e.g. lung cancer -  
 XX Claim 1; Page 334; 378pp; English.

XX The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer. The  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is a cDNA encoding human lung tumour-specific protein.

XX Sequence 337 BP; 103 A; 60 C; 93 G; 81 T; 0 other;

Alignment Scores:  
 Pred. No.: 5.82e-54 Length: 337  
 Score: 532.00 Matches: 97  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0

US-09-854-133-586 (1-97) x AAD23462 (1-337)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 Db 5 GAGGTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTCTGTCTCAAACA 64  
 QY 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
 Db 65 GAATTAGGAAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAATCCAGGCCAATTGTGGC 124  
 QY 41 IleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
 Db 125 ATAGATTTTATCATATCTGGATTTTGGATTTCTTTTCTCATCTGATTCAG 184  
 QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 Db 185 GAAAGCCTGTGTGTCCACCATCTCCAAAGGAGGTACCTGCGAGGAAATGTTAAGCGGA 244  
 QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgGlyCysSer 97  
 Db 245 GCCTGCCTTCCTGGCAACAGGAGCCACCTGGCGAGAGAAAGTGCAGC 295

RESULT 2

AAD23461  
 ID AAD23461 standard; cDNA; 5981 BP.

XX AAD23461;

XX 26-FEB-2002 (first entry)

XX Human lung tumour-specific 14F10 full length cDNA.

XX

KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 KW antisense-therapy; vaccine; immune response; lung cancer; 14F10; ss.

XX Homo sapiens.

XX WO200172295-A2.

XX 04-OCT-2001.

XX 28-MAR-2001; 2001WO-US09991.

XX 29-MAR-2000; 2000US-0538037.

PR 05-JUN-2000; 2000US-0588937.

PR 18-AUG-2000; 2000US-0640878.

PR 22-SEP-2000; 2000US-234517P.

PR 01-NOV-2000; 2000US-0704512.

PR 14-DEC-2000; 2000US-0738973.

XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Mohamath R, Secrlist H, Benson DR, Indirias CY;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX WPI; 2001-639201/73.

XX New human lung-specific polynucleotides and polypeptides for the  
 PT diagnosis and treatment of disease e.g. lung cancer -  
 XX Claim 1; Page 332-334; 378pp; English.

XX The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer. The  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is a cDNA encoding human lung tumour-specific protein.

XX Sequence 5981 BP; 1858 A; 1029 C; 1098 G; 1996 T; 0 other;

Alignment Scores:

Pred. No.: 1.18e-51 Length: 5981  
 Score: 527.00 Matches: 96  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.06% Indels: 0  
 DB: 23 Gaps: 0

US-09-854-133-586 (1-97) x AAD23461 (1-5981)

QY 2 ValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGlu 21  
 Db 3 GTTGAAGTGAGCAGAGATCATGCCAGCCTGGGTGACAGTCTGTCTCAAACAGAA 62  
 QY 22 LeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGlyIle 41  
 Db 63 TTAAGGAAAAAAGAAAGAAAGAAAGAGAGAGAGAAATCCAGGCCAATTGTGCATA 122  
 QY 42 AspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGlnGlu 61  
 Db 123 GATTTTATCATATCTGGATTTTGGATTTCTTTTCTCATCTGATTCAGGAA 182

QY 62 SerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGlyGly 81  
 Db 183 AGCCTGTGTGTCCACCATCTCCAAAGAGAGGTACCTGCGAGGAAATGTTAAGCGAGGC 242

QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgGlyCysSer 97

Db 243 TGCCTTCCTGGCAACAGGAGGCCACCTGGCGAGAGAAAGTGCAGC 290

## RESULT 3

AAD23460  
 ID AAD23460 standard; cDNA; 2239 BP.  
 XX  
 AC AAD23460;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE Human lung tumour-specific 19A4 cDNA.  
 KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 KW antisense-therapy; vaccine; immune response; lung cancer; 19A4; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200172295-A2.  
 XX  
 PD 04-OCT-2001.  
 XX  
 PF 28-MAR-2001; 2001WO-US09991.  
 XX  
 PR 29-MAR-2000; 2000US-0538037.  
 PR 05-JUN-2000; 2000US-0588937.  
 PR 18-AUG-2000; 2000US-0640878.  
 PR 22-SEP-2000; 2000US-234517P.  
 PR 01-NOV-2000; 2000US-0704512.  
 PR 14-DEC-2000; 2000US-0738973.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Reed SG, Lodes MJ, Mohamath R, Secretist H, Benson DR, Indirias CY;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX  
 DR WPI; 2001-639201/73.  
 XX  
 PT New human lung-specific polynucleotides and polypeptides for the  
 PT diagnosis and treatment of disease e.g. lung cancer -  
 XX  
 PS Claim 1; Page 332; 378pp; English.  
 XX  
 CC The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer. The  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is a cDNA encoding human lung tumour-specific protein.  
 XX  
 SQ Sequence 2239 BP; 619 A; 444 C; 493 G; 683 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.39e-50 Length: 2239  
 Score: 513.00 Matches: 97  
 Percent Similarity: 84.35% Conservative: 0  
 Best Local Similarity: 84.35% Mismatches: 18  
 Query Match: 96.43% Indels: 1  
 DB: 23 Gaps: 1  
 US-09-854-133-586 (1-97) x AAD23460 (1-2239)  
 QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 Db 2 GAGGTGAAGTGAAGCAGAGATCATGCCACCCCTGGGTGACAGTCTGTCTCAACA 61  
 QY 21 GluLeuArgLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
 Db 62 GAATTAAGGAAAAAGAAAAAGAAAAAGAGAGAGAGAGAAATCCAGGCCAATTTGTGGC 121  
 QY 41 IleAspPheIlePheTrpIlePheTrpIleLeuPheSerHisHisTrpIleGln 60  
 Db 122 ATAGATTTTATCATATTTCTGGATTTTGGATTCTTTTGTTCATCATCTGATTCAG 181

QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 Db 182 GAAGCCCTGTTGTGTCACCATCTCCAAAGGAGGTTACCTGCAGGAAATGTTACGGGA 241  
 QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg----- 93  
 Db 242 GGCTGCTTCCCTGGGCAACAAGAGCCACCTGGCAGGACGCTTTTCAGGAAGAGACG 301  
 QY 94 -----ArgLysCysSer 97  
 Db 302 CCTTTTCAGGAAGAGAGCGCTTTTCAGGAAGAGAAAGTGCAGC 346

## RESULT 4

ABL37718  
 ID ABL37718 standard; cDNA; 575 BP.  
 XX  
 AC ABL37718;  
 XX  
 DT 08-APR-2002 (first entry)  
 XX  
 DE Human colon tumour antigen polynucleotide SEQ ID NO:1307.  
 XX  
 KW Human; colon cancer; colon tumour antigen; cytostatic; vaccine;  
 KW colon tumour metastatic antigen; diagnosis; gene; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200196388-A2.  
 XX  
 PD 20-DEC-2001.  
 XX  
 PF 08-JUN-2001; 2001WO-US18557.  
 XX  
 PR 09-JUN-2000; 2000US-210899P.  
 PR 20-FEB-2001; 2001US-270216P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Jiang Y, Harlocker SL, Secretist H;  
 XX  
 DR WPI; 2002-114514/15.  
 XX  
 PT Novel isolated colon tumor polynucleotide differentially expressed in  
 PT colon tumor or colon metastatic tumor and polypeptides encoded by them,  
 PT useful for inhibiting development of cancer in patient -  
 XX  
 PS Claim 1; SEQ ID 1307; 105pp; English.  
 XX  
 CC ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)  
 CC which were isolated from human colon tumour and colon metastatic tumour  
 CC cDNA libraries. (I) have cytostatic activity and can be used in vaccine  
 CC production. (I) can be used for stimulating and/or expanding T cells  
 CC specific for a tumour protein on contact with the T cells. They are also  
 CC useful for inhibiting the development of cancer in a patient. (I) can be  
 CC used as probes or primers for nucleic acid hybridisation, for preparing  
 CC mutant species primers, or primers for use in genetic constructions. (I)  
 CC can be used in the diagnosis of a colon tumour.  
 XX  
 SQ Sequence 575 BP; 151 A; 122 C; 153 G; 141 T; 8 other;  
 Alignment Scores:  
 Pred. No.: 8.13e-17 Length: 575  
 Score: 221.00 Matches: 47  
 Percent Similarity: 70.00% Conservative: 9  
 Best Local Similarity: 58.75% Mismatches: 18  
 Query Match: 41.54% Indels: 6  
 DB: 24 Gaps: 2  
 US-09-854-133-586 (1-97) x ABL37718 (1-575)  
 QY 19 GlnThrGluLeuArgLysLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37

Db 142 CAGTCTGAAGCAGGAGACATCGATCAGTAACACCAAGAGACACCAAAAGTTGAAAGT 201  
 QY 38 AsnCysGlyIleAspPheIlePheTrpIleLeuLeuPheSerHis 57  
 Db 202 TTGTGTTCTTCCTCTCTGTTTATTTTCCCGGTGTGCTTACTATGTCA----- 255  
 QY 58 TrpIleGlnGluSerProSerProSerProLysGluValThrCysArgGluMet 77  
 Db 256 -----GAAAGCCCTGTGTGTCCACCATCTCCAAAGAGAGTTACCTGCAGGAAATG 306  
 QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyValArgGlyCysSer 97  
 Db 307 TTAACGGGAGGCTCCCTCCCTGGGCAACAGAGACCCACCTGGCGCAGGAGAAAGTGCAGC 366

## RESULT 5

ID ABK27651/C  
 ID ABK27651 standard; cDNA; 596 BP.

AC ABK27651;

DT 09-APR-2002 (first entry)

DE Human colon cancer expressed sequence tag, Seq ID no 88.

KW Human; colon cancer; T cell expansion; tumour; EST; gene; ss;

KW expressed sequence tag.

OS Homo sapiens.

PN WO200196390-A2.

XX 20-DEC-2001.

PF 08-JUN-2001; 2001WO-US18577.

PR 09-JUN-2000; 2000US-210821P.

PR 18-DEC-2000; 2000US-256571P.

PR 10-MAY-2001; 2001US-290240P.

XX (CORI-) CORIXA CORP.

XX Jiang Y, Hepler WT, Clapper JD, Wang A, Secrist H;

XX WPI; 2002-139708/18.

XX Novel isolated polynucleotide encoding a polypeptide comprising a

XX portion of colon tumour protein, useful for detection, diagnosis and

XX therapy of human colon cancer

XX Claim 1; Page 174; 220pp; English.

CC The invention relates to an isolated polynucleotide (I) encoding a  
 CC polypeptide (II) comprising at least a portion of a colon tumour  
 CC protein. (I), (II) and antibody (III) to (II) are useful for determining  
 CC the presence of a cancer in a patient. (I), (II) or antigen presenting  
 CC cells expressing (I) is useful for stimulating and/or expanding T cells  
 CC specific for a tumour protein, by contacting T cells with (I), (II) or  
 CC antigen-presenting cells that express (I), under conditions and for a  
 CC time sufficient to permit the stimulation and/or expansion of T cells.  
 CC (I), (II), or antigen presenting cells that express (II) are useful for  
 CC treating colon cancer in a patient by incubating CD4+ and/or CD8+ T cells  
 CC isolated from a patient with (I), (II) or antigen presenting cells that  
 CC express (II), such that T cells proliferate, and administering to the  
 CC patient an effective amount of the proliferated T cells, thus inhibiting  
 CC the development of a cancer in the patient. (I) or (II) is useful in  
 CC vaccines and pharmaceutical compositions for prevention and treatment  
 CC of colon malignancies and for the diagnosis and monitoring of such  
 CC cancers. (I), (II) or (III) is useful for detection, diagnosis and/or  
 CC therapy of human colon cancer. (I) is useful as a probe or primer for  
 CC nucleic acid hybridisation, and in the design and preparation of  
 CC ribozyme molecules for inhibiting expression of (II) in tumour cells.  
 CC ABK27564-ABK27807 represent novel human colon cancer coding  
 CC sequences and primers of the invention.

XX

SQ Sequence 596 BP; 160 A; 151 C; 126 G; 158 T; 1 other;

Alignment Scores:

Pred. No.: 8.54e-17 Length: 596

Score: 221.00 Matches: 47

Percent Similarity: 70.00% Conservative: 9

Best Local Similarity: 58.75% Mismatches: 18

Query Match: 41.54% Indels: 6

Db: 24 Gaps: 2

US-09-854-133-586 (1-97) x ABK27651 (1-596)

QY 19 GlnThrGluLeuArgLysLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37

Db 481 CAGTCTGAAGCAGGAGACATCAGTAACACCAAGAGACACCAAAAGTTGAAAGT 422

QY 38 AsnCysGlyIleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHis 57

Db 421 TTGTGTTCTTCCTCTCTGTTTATTTTCCCGGTGTGCTTACTATGTCA----- 368

QY 58 TrpIleGlnGluSerLeuLeuCysProSerProLysGluValThrCysArgGluMet 77

Db 367 -----GAAAGCCCTGTGTGTCCACCATCTCCAAAGAGAGTTACCTGCAGGAAATG 317

QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyValArgGlyCysSer 97

Db 316 TTAACGGGAGGCTCCCTCCCTGGGCAACAGAGACCCACCTGGCGCAGGAGAAAGTGCAGC 257

## RESULT 6

AAZ16609

ID AAZ16609 standard; cDNA; 800 BP.

XX AAZ16609;

DT 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:4079.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

XX detection; mapping; tissue typing; profiling; forensic; cancer;

XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

XX WO9938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

XX 28-JAN-1998; 98US-0072910.

XX 24-FEB-1998; 98US-0075954.

XX 31-MAR-1998; 98US-0080114.

XX 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX CrkVenjakov R, Dickson M, Drmanac R, Drmanac S;

XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

XX Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

XX differentially expressed in different cell types

XX Claim 1; Page 1934; 2479pp; English.



\_\_\_\_\_

---



Oy 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCys 96  
Db 65 TTAACGGAGGCTGCTTCCCTGGGCAACAAGAGGCCACCTGGGCAAGAAAGTGC 121  
RESULT 10  
AAL04753  
ID AAL04753 standard; DNA; 18408 BP.  
XX AAL04753;  
AC AAL04753;  
XX 21-NOV-2001 (first entry)  
DT Human reproductive system related antigen DNA SEQ ID NO: 7441.  
DE Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
KW Homo sapiens.  
XX WO200155320-A2.  
XX 02-AUG-2001.  
PD 17-JAN-2001; 2001WO-US01339.  
PF 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217486.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232403.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.









us-09-854-133-586\_1.rng

PR	17-MAR-2000;	2000US-0190076.
PR	18-APR-2000;	2000US-0198123.
PR	19-MAY-2000;	2000US-0205515.
PR	27-JUN-2000;	2000US-0209467.
PR	28-JUN-2000;	2000US-0214886.
PR	30-JUN-2000;	2000US-0215135.
PR	07-JUL-2000;	2000US-0216647.
PR	07-JUL-2000;	2000US-0216880.
PR	11-JUL-2000;	2000US-0217487.
PR	11-JUL-2000;	2000US-0217496.
PR	16-JUL-2000;	2000US-0218290.
PR	26-JUL-2000;	2000US-0220963.
PR	26-JUL-2000;	2000US-0220964.
PR	14-AUG-2000;	2000US-0224518.
PR	14-AUG-2000;	2000US-0224519.
PR	14-AUG-2000;	2000US-0225214.
PR	14-AUG-2000;	2000US-0225213.
PR	14-AUG-2000;	2000US-0225266.
PR	14-AUG-2000;	2000US-0225267.
PR	14-AUG-2000;	2000US-0225268.
PR	14-AUG-2000;	2000US-0225270.
PR	14-AUG-2000;	2000US-0225447.
PR	14-AUG-2000;	2000US-0225757.
PR	14-AUG-2000;	2000US-0225758.
PR	18-AUG-2000;	2000US-0225759.
PR	18-AUG-2000;	2000US-0226279.
PR	22-AUG-2000;	2000US-0226681.
PR	22-AUG-2000;	2000US-0226688.
PR	22-AUG-2000;	2000US-0227182.
PR	23-AUG-2000;	2000US-0227009.
PR	30-AUG-2000;	2000US-0228924.
PR	01-SEP-2000;	2000US-0229287.
PR	01-SEP-2000;	2000US-0229343.
PR	01-SEP-2000;	2000US-0229344.
PR	05-SEP-2000;	2000US-0229345.
PR	05-SEP-2000;	2000US-0229509.
PR	06-SEP-2000;	2000US-0229513.
PR	06-SEP-2000;	2000US-0230437.
PR	08-SEP-2000;	2000US-0231242.
PR	08-SEP-2000;	2000US-0231243.
PR	08-SEP-2000;	2000US-0231244.
PR	08-SEP-2000;	2000US-0231413.
PR	08-SEP-2000;	2000US-0231414.
PR	08-SEP-2000;	2000US-0232080.
PR	12-SEP-2000;	2000US-0232081.
PR	14-SEP-2000;	2000US-0231968.
PR	14-SEP-2000;	2000US-0232397.
PR	14-SEP-2000;	2000US-0232398.
PR	14-SEP-2000;	2000US-0232399.
PR	14-SEP-2000;	2000US-0232400.
PR	14-SEP-2000;	2000US-0232401.
PR	14-SEP-2000;	2000US-0233063.
PR	14-SEP-2000;	2000US-0233064.
PR	21-SEP-2000;	2000US-0233065.
PR	21-SEP-2000;	2000US-0234223.
PR	21-SEP-2000;	2000US-0234274.
PR	25-SEP-2000;	2000US-0234997.
PR	25-SEP-2000;	2000US-0234998.
PR	26-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	27-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	02-OCT-2000;	2000US-0236670.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	13-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.

Alignment Scores:

Pred. NO.:	0.591	Length:	18408
Score:	104.50	Matches:	32
Percent Similarity:	53.91%	Conservative:	16
Best Local Similarity:	35.9%	Mismatches:	23
Query Match:	19.64%	Indels:	18
DB:	23	Gaps:	4

Y		2	ValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrIleuSerGlnThrGlu	21
b		15928	GTGGAGATCGCACCACCTGCACCTCCAGCCCTGGTGATAGCAAAACTCTCTCAAAA---	15984
Y		22	LeuArgLysGluArgLysLysArgGluArgLysPheGlnAlaAsnCysGlyTyrIle	41
b		15985	---AAGAAAAAGAGAAGAGAAAAATAATCACCCCTTTATT---	16026
Y		42	AspPheIlePheTrpIlePheTrpIleLeuPheSerHis-----His	57
b		16027	TATTTTTTCCTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTCATCTGTCGCCAGGC	16086
Y		58	Trp-IleGlnGluSerLeuLeuCys-----ProProSerProLysGlu	71
b		16087	TGGAGTGAACAATGGCCGATCTCTGCTACTACAAACCTCCACCTCCCGGGTCAAGCG	16146
Y		71	uValThrCysArgGluMetLeuThr	79
b		16147	ATTCTTCCTGCGCTCAGCCCTCTGAGT	16171

ESQUL 13  
8L97657

ABI.97657.

21-JUN-2002 (first entry)

Human testicular antigen encoding DNA fragment	SEQ ID NO.
Human testicular antigen encoding DNA fragment	SEQ ID NO. 3300

**Homo sapiens.**

WO200155317-A2.

02-AUG-2001.

1. SAN-2001; 2001WO-US01329.

04-FEB-2000; 2000US-0180628,

02-MAR-2000; 2000US-0186350.  
16-MAR-2000; 2000US-0186350.



CC The present invention is concerned with the use of the human tumour  
 CC suppressor gene BRG1 in cancer diagnosis and therapy. This gene is  
 CC comprised of several exons, shown in AAC58874-C58903, and has several  
 CC splice variants, given in AAC58906-C58912. The protein sequences for  
 CC these are shown in AAB27552-B27558. BRG1 is a homologue of the Drosophila  
 CC protein brhma, and has been shown to be bound to retinoblastoma tumour  
 CC suppressor protein RB. The BRG1 coding sequence and protein can be used  
 CC in the diagnosis and treatment of cancer (for example by gene therapy),  
 CC particularly prostate cancer, to identify drugs useful in the treatment  
 CC of cancer and in the production of animal models for cancer.

SQ Sequence 971 BP; 229 A; 226 C; 266 G; 250 T; 0 other;

# Alignment Scores:

Pred. No.:	0.0206	Length:	971
Score:	102.00	Matches:	33
Percent Similarity:	45.92%	Conservative:	12
Best Local Similarity:	33.67%	Mismatches:	28
Query Match:	19.17%	Indels:	25
DB:	21	Gaps:	6

US-09-854-133-586 (1-97) x AAC58883 (1-971)

QY	3	GluValSerArgAspHisAlaSerIeuGlyAspSerGluThrLeuSerGlnThrGluLeu	22
Db	22	GAGATACGCCACGCTAGCTGGTGAC-----AGACCGAGACTC	66
QY	23	Arg---LysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGlyIle	41
Db	67	CGTCTCAAAAGAAAAAAGAGAAAGAACAGTGTCAAGGTTG---ATG	123
QY	42	AspPheIle-IlePheTrpIle-----PheTrpIleLeuLeuPheSerHisHistrpI	59
Db	124	GGTTATATATTTTACTGGATTACTAGTATACAGTTTACGATTTTCCACCACCTGG--	181
QY	59	eGlnGluSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu	79
Db	182	-----CCAGTTGGCTGCCATTCTGTGCAACAGTTTCAAGT	219
QY	79	rcGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCys	96
Db	220	GGGTGCT-----TCCACCTTGGCTCTGTAAGTGT	250

# RESULT 15

AAL06199/c

ID AAL06199 standard; DNA; 3424 BP.

AC AAL06199;

XX

DT 21-NOV-2001 (first entry)

XX

DE Human reproductive system related antigen DNA SEQ ID NO: 8887.

XX Human; reproductive system related antigen; reproductive system disorder;  
 KW cancer; gene therapy; ds.

XX Homo sapiens.

OS

XX WO200155320-A2.

PN

XX 02-AUG-2001.

PD

XX 17-JAN-2001; 2001WO-US01339.

PF

XX 31-JAN-2000; 2000US-0179065.

PR

XX 04-FEB-2000; 2000US-0180628.

PR

XX 24-FEB-2000; 2000US-0184664.

PR

XX 02-MAR-2000; 2000US-0186350.

PR

XX 16-MAR-2000; 2000US-0189874.

PR

XX 17-MAR-2000; 2000US-0190076.

PR

XX 18-APR-2000; 2000US-0196123.

PR

XX 19-MAY-2000; 2000US-0205515.

PR

XX 07-JUN-2000; 2000US-0209467.

PR

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX









```
QY 23 Arg---LysLysGluArgLysLysArgGluArgLysPheGlnAlaAsnCysGlyIle 41
Db 67 CQCTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGTTG---ATG 123
QY 42 AspPheIle-IlePheTrpIle-----PheTrpIleLeuLeuPheSerHisHisTrpIle 59
Db 124 GGTATATATATTTTACTGGATTAAGTACTAGTATACAGATTACGCAATTTCCACCACCTGG-- 181
QY 59 eGlnGluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuTh 79
Db 182 -----CCATGTTTGGCTGCCATTTTCTGTGCAACACATTGAGT 219
QY 79 rGlyGlyCysLeuProTrpAlaThrArgSerHisLeuLeuArgArgLysCys 96
Db 220 GGGTGGCT-----TCCACCTTGGCTCTGTAAAGTGT 250

RESULT 2
US-09-798-096-10/c
; Sequence 10, Application US/09798096
; Patent No. 6399378
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF RECQL2 EXPRESSION
; FILE REFERENCE: RFS-0207
; CURRENT APPLICATION NUMBER: US/09/798,096
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 10
; LENGTH: 99500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-798-096-10

Alignment Scores:
Pred. No.: 38.4 Length: 99500
Score: 91.50 Matches: 39
Percent Similarity: 38.35% Conservative: 12
Best Local Similarity: 29.32% Mismatches: 21
Query Match: 17.20% Indels: 61
DB: 4 Gaps: 9

US-09-854-133-586 (1-97) x US-09-798-096-10 (1-99500)
QY 8 HisLaserLeuGlyAsp---SerGluThrLeuSerGlnThrGluLeuArgLysGlu 26
Db 43428 CACTCCAGCCTGGGGACAAAAAGTCAGACTTTGTCT-----CACAAAAAAG 43381
QY 27 ArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGlyIleAspPheIlePhe 46
Db 43380 AAGAAGAAGAAGAAGAAGAAG-----GGGGTC----- 43351
QY 47 TrpIlePheTrpIle-----LeuLeuPhe--- 54
Db 43350 -----TGGCTCTGTGGCCAGGTAAACTGCAGTGGTGAATCACTGCTCACTGAA 43300
QY 55 ---SerHisHisTrpIleGlnGluSerLeuLeu---CysProSerProLysGluVal 72
Db 43299 CCCTCAAACTCTTGGGCTCAAGAGATCTCTGCTGCCTCAGCCCTCCCAAGTAGTGAGACT 43240
QY 73 Thr----- 73
Db 43239 ACAGGCCTATGCCACTGTGCCCAACACACATCTCTCTCTTTTCTTTTAA 43180
QY 74 -----CysArgGluMetLeuThrGlyGlyCysLeuProTrpAlaThr----- 87
Db 43179 TTAATATATTTTCTAAGACAGGCTGTGCGATGTTGCCAGGCTGGTCTCAAACTCCT 43120
QY 88 -----ArgSerHisLeuGlyArgArgLysCys 96
Db 43119 GGGCTCAAGTGATCTCCCGCCTTGGGCTCTCTAAAGTGC 43081
```

```
RESULT 3
US-08-406-030A-4/c
; Sequence 4, Application US/08406030A
; Patent No. 6270989
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Haug, Brian M.
; APPLICANT: Selden, Richard F.
; TITLE OF INVENTION: Protein Production and Delivery
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/406,030A
; FILING DATE: 17-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/243,391
; FILING DATE: 13-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,586
; FILING DATE: 03-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/911,533
; FILING DATE: 10-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,840
; FILING DATE: 05-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/789,188
; FILING DATE: 05-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11704
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09627
; FILING DATE: 05-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: TKT95-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-406-030A-4

Alignment Scores:
Pred. No.: 1.2 Length: 2455
Score: 85.50 Matches: 34
Percent Similarity: 48.04% Conservative: 15
Best Local Similarity: 33.33% Mismatches: 27
Query Match: 16.07% Indels: 26
DB: 4 Gaps: 6
```

```
US-09-854-133-586 (1-97) x US-08-406-030A-4 (1-2455)
QY 3 GluValSerArgAspHisAlaSerLeuGlyAsp----SerGluThrLeuSerGlnThrGlu 21
Db 1352 GAGATCGACCATTCAGCTGCGCTGAGCAATAGAGTGAACCTCCGCTCTCAAAAAAA 1293
QY 22 LeuArgLysLysGluArgGluArgLysPheGlnAlaAsnCysGlyIle 41
Db 1292 AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1245
QY 42 AspPheIleIlePheTrpIlePheTrpIleLeuPheSerHisHisTrpIleGlnGlu 61
Db 1244 TCCTTTTGTGAGG-----CTCCTCAGTGAA-----TCGGAC 1212
QY 62 SerLeuLeuCysProSerProLysGlu-ValThrCysArgGluMetLeuThr----- 79
Db 1211 ATGATTGGTGCCACCTACCCCTGCTGAGAGCAGCAGACATGATATATACACACAT 1152
QY 80 -----GlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgAr 94
Db 1151 ATCTTAATTAGCCTTAATTTCTGAGCAACCTTGGTTG-----TTAGGTGAGAG 1104
QY 94 gLys 95
Db 1103 AAAA 1100

RESULT 4
US-07-906-871-15
; Sequence 15, Application US/07906871
; Patent No. 5340739
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Avraham, Shalom
; TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W., Suite 300
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/906,871
; FILING DATE: 19920103
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/816,289
; FILING DATE: 03 JAN 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/635,544
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US89/03051
; FILING DATE: 13-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/224,035
; FILING DATE: 13-JUL-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 0627.2830004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)833-7533
; TELEFAX: (202)833-8716
; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 17327 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: exon
; LOCATION: 621..753
; FEATURE:
; NAME/KEY: intron
; LOCATION: 754..9596
; FEATURE:
; NAME/KEY: exon
; LOCATION: 9597..9744
; FEATURE:
; NAME/KEY: intron
; LOCATION: 9745..16396
; FEATURE:
; NAME/KEY: exon
; LOCATION: 16397..17327
; US-07-906-871-15

Alignment Scores:
Pred. No.: 23.5 Length: 17327
Score: 84.50 Matches: 24
Percent Similarity: 54.67% Conservative: 17
Best Local Similarity: 32.00% Mismatches: 11
Query Match: 15.88% Indels: 23
DB: 1 Gaps: 5

US-09-854-133-586 (1-97) x US-07-906-871-15 (1-17327)
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAsp-----SerGluThr 16
Db 4329 GAGTTGCGAGTGAGCCAGATCATGCCACTCCAGCTCCAGCCGCGGATAGACTGAGACT 4388
QY 17 LeuSerGlnThrGluLeuArgLysLysGluArgLysLysLysLysLysLysLysLysLys 36
Db 4389 CTGCTCAAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4445
QY 37 AlaAsnCys-----GlyIleAspPhe 43
Db 4446 GTATATTGTGATTATTGTTGAGGAGAAAGTAGTACCATATATAAAGGTATGGAC--- 4502
QY 44 IleIlePheTrp-----IlePheTrpIleLeuLeuPheSerHis 56
Db 4503 -----TATTGAGAGAAAGTTGTTGCTTGGTAAACATTACTCAT 4541

RESULT 5
US-08-869-696-1/c
; Sequence 1, Application US/08869696C
; Patent No. 6031155
; GENERAL INFORMATION:
; APPLICANT: Cameron-Mills, Verena
; APPLICANT: Lok, Finn
; APPLICANT: Sinjorgo, Catharina Maria Cornelia
; APPLICANT: Van Den Dool, Ronald Tako Marinus
; APPLICANT: Caspers, Martinus Petrus Maria
; APPLICANT: Van Zelfj-van Der Valk, Maria Joanna
; TITLE OF INVENTION: ARABINOXYLAN DEGRADATION
; FILE REFERENCE: 11225.01US01
; CURRENT APPLICATION NUMBER: US/08/869,696C
; CURRENT FILING DATE: 1997-06-05
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5529
; TYPE: DNA
; ORGANISM: barley
; US-08-869-696-1

Alignment Scores:
```



```

MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT-US95/07201
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/367,841
FILING DATE: 30-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,963
FILING DATE: 07-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126PFI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 22481 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: PI-147
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: full length gene
OTHER INFORMATION: sequence for P1
PCT-US95-07201-43
Alignment Scores:
Pred. No.: 38.7
Score: 84.00
Percent Similarity: 34.92%
Best Local Similarity: 23.02%
Query Match: 15.79%
DB: 5
US-09-854-133-586 (1-97) x PCT-US95-07201-43
QY 3 GluValSerArgAspHisAlaSerLeuGlyVal
| | | | | | | | | | | | | | | | | |
Ddb 13050 GAGTCGGCAGCCACTGCACCTCAGCGTGGCG
QY 23 Arg---LysLysGluArgLysLysLysArgGG
| | | | | | | | | | | | | | | | | |
Ddb 13095 CGTCTCAAAAAAAAAAAAAAAGAAAG
QY 36 -----
Ddb 13155 TTCCTTTTTTTAGACAGAGTCTCACCTCCAT
QY 37 ---AlaAsnGlyGlyIleAspPhe-----
| | | | | | | | | | | | | | | | | |
Ddb 13215 TTGGGTCACTGAATCTTTGGCTCCTCGATT
QY 44 -----
Ddb 13275 ATAGCTGGGACACACAGGCACCTGCCACACG
QY 52 LeuLeuPheSerHisHisTrpIleGlnclu
| | | | | | | | | | | | | | | | | |

```

## RESULT 9

US-08-965-729A-2  
; Sequence 2, Application US/08965729A  
; Patent No. 6200751  
; GENERAL INFORMATION:  
; APPLICANT: Jian-Ming Gu and Charles T. Esmon  
; TITLE OF INVENTION: ENDOTHELIUM SPECIFIC EXPRESSION  
; TITLE OF INVENTION: REGULATED BY EPCR CONTROL ELEMENTS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center, 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-4530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/965.729A  
; FILING DATE: 07-NOV-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/030,718  
; FILING DATE: 08-NOV-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: OMRF 164 PCT  
; TELEPHONE: 404-873-8794  
; TELEFAX: 404-873-8795  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3224 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: /note= "Nucleotides 2272 through 2729 are a  
; OTHER INFORMATION: thrombin responsive element"; Human  
US-08-965-729A-2

## Alignment Scores:

Pred. No.:	3.03	Length:	3224
Score:	83.50	Matches:	30
Percent Similarity:	50.62%	Conservative:	11
Best Local Similarity:	37.04%	Mismatches:	15
Query Match:	15.70%	Indels:	25
DB:	4	Gaps:	6

US-09-854-133-586 (1-97) x US-08-965-729A-2 (1-3224)

QY	1	GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr	20
Db	982	GAGGTGACGTGAGCCGAGATCAGCCGACCTGCAFTACAGCAAGACTCCATCTCA	1035
QY	21	GluLeuArgLysLysGluArgLysLysArgGluArgLysPheGlnAlaAsnCysGly	40
Db	1036	-----AAAAAGAAAAAAGAAAAAGAA-----TGT---	1068
QY	41	IleAspPheIleIlePhe-----Trp-----IlePheTrpIle	51
Db	1069	-----TTCAATAATTTTAAATAAGGCAAGACAATAATAATTTGGTATTATTAACTCA	1122
QY	52	LeuLeuPheSerHisHisTrpIleGlnGluSer-----LeuLeuCysProProSer	68

Db 1123 TTCCTACTTTTCTGAGGCCAGTCGAGGAAACAAAGTTCCTCTCTGTTCCAACTAGA 1182  
QY 69 Pro 69  
Db 1183 CCA 1185  
RESULT 10  
US-09-053-866-1/c  
; Sequence 1, Application US/09053866  
; Patent No. 6111075  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Wenfeng  
; APPLICANT: Presnell, Scott R.  
; APPLICANT: Yee, David P.  
; APPLICANT: Foster, Donald C.  
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/053,866  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leith, Debra K  
; REGISTRATION NUMBER: 32,619  
; REFERENCE/DOCKET NUMBER: 98-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6674  
; TELEFAX: 206-442-6678  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4895 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 176...1330  
; OTHER INFORMATION:  
US-09-053-866-1

Alignment Scores:	5.4	Length:	4895
Pred. No.:	83.50	Matches:	35
Score:	41.74%	Conservative:	13
Best Local Similarity:	30.43%	Mismatches:	37
Query Match:	15.70%	Indels:	31
DB:	3	Gaps:	6

US-09-854-133-586 (1-97) x US-09-053-866-1 (1-4895)

QY	1	GluValGluValSerArgAspHisAlaSerLeu-----GlyAspSerGluThr	16
Db	4362	GAGCTTGACGTGAGCCGAGATCAGCCACTGCAGCTCCAGCTGGGCAACAGCCGAGACT	4303
QY	17	LeuSerGlnThrGluLeuArgLysLysGluArgLysLysArgGluArgLysPheGln	36



```

; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5377.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; OTHER INFORMATION: The sequence is preceded by an
; unsequenced portion of from 4.7 to 5.3
; OTHER INFORMATION: kilobases
US-09-854-133-586 (1-97) x US-09-814-951A-3 (1-9704)

Alignment Scores:
Pred. No.: 4.89 Length: 4129
Score: 83.00 Matches: 28
Percent Similarity: 51.35% Conservative: 10
Best Local Similarity: 37.84% Mismatches: 12
Query Match: 15.60% Indels: 24
DB: 4 Gaps: 4

US-09-854-133-586 (1-97) x US-09-224-834-12 (1-4129)
QY 1 GluValGluValSerArgAspHisAlaSerLeu-----GlyAspSerG
Db 1441 GAGGTTCCAGTCAGCCGAGATCGTCCATTATATCTCCAGCCTGGGCAACAGAGTGG
QY 17 LeuSerGlnThrGluLeuArgLysLysLysGluArgLysLysLysLysArgGluArgLysP
Db 1501 CTGCTCAA-----AAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT
QY 37 AlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPhe
Db 1549 -----TTATTTACCTTCACATTATCTTC
QY 55 -----SerHisHisTrpIleGlnGluSerLeuLeuCys 65
Db 1576 GGATGTTCTTCCTTTATGTAGGTACCAAGGTTCTGCACCTATGT 1617

RESULT 14
US-09-814-951A-3
; Sequence 3, Application US/09814951A
; Patent No. 6387661
; GENERAL INFORMATION:
; APPLICANT: SHAO, Wei et al
; TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
; FILE OF INVENTION: ACID MOLECULES ENCODING HUMAN AMINOACYLASE
; FILE REFERENCE: CL001179
; CURRENT APPLICATION NUMBER: US/09/814,951A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9704
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-951A-3

Alignment Scores:
Pred. No.: 18.3 Length: 9704
Score: 82.50 Matches: 19
Percent Similarity: 74.42% Conservative: 13
Best Local Similarity: 44.19% Mismatches: 8
Query Match: 15.51% Indels: 3
DB: 4 Gaps: 2

US-09-854-133-586 (1-97) x US-09-814-951A-3 (1-9704)

```







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 15:28:15 ; Search time 115.027 Seconds  
(without alignments)  
1047.953 Million cell updates/sec

Title: US-09-854-133-586

Perfect score: 532

Sequence: 1 EVEVSRDHASIGDSETLSQT.....LTGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh  
-MODEL=frame\_p2n.model -US09854133/runat\_05052003\_173958\_449/app\_query.fasta\_1.462  
-DB=/cn2\_1/USPTO.spool/US09854133/runat\_05052003\_173958\_449/app\_query.fasta\_1.462  
-Q=Published Applications NA -QPM=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09854133\_@CGN\_1\_1\_117\_@runat\_05052003\_173958\_449  
-NCPU=6 -ICPU=3 -NO\_XLPHY -NO\_MMAP -TIMEOUT=30 -THREADS=1 -XGAPEXT=0.5 -FGAPOPEXT=6  
-DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7  
-FGAPEXT=7 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
10: /cn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
11: /cn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
12: /cn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:  
13: /cn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
14: /cn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	532	100.0	337	9 US-09-854-133-442	Sequence 442, App
2	532	100.0	337	10 US-09-738-973-442	Sequence 442, App
3	527	99.1	5981	9 US-09-854-133-441	Sequence 441, App
4	527	99.1	5981	10 US-09-738-973-441	Sequence 441, App

5	513	96.4	2239	9 US-09-854-133-440	Sequence 440, App
6	513	96.4	2239	10 US-09-738-973-440	Sequence 440, App
7	221	41.5	572	9 US-10-046-935-1307	Sequence 1307, App
8	221	41.5	572	9 US-09-878-178-1307	Sequence 1307, App
9	221	41.5	572	9 US-10-146-502-1307	Sequence 1307, App
10	221	41.5	1861	9 US-10-163-866-30	Sequence 30, Appl
11	221	41.5	2000	9 US-10-163-866-34	Sequence 34, Appl
12	221	41.5	2482	9 US-10-163-866-29	Sequence 29, Appl
13	210.5	33.6	520	9 US-10-163-866-32	Sequence 32, Appl
14	210.5	33.6	1268	9 US-10-163-866-53	Sequence 53, Appl
15	210.5	33.6	1528	9 US-10-163-866-52	Sequence 52, Appl
16	210.5	33.6	1542	9 US-10-163-866-33	Sequence 33, Appl
17	210.5	39.6	3144	9 US-10-163-866-31	Sequence 31, Appl
18	104.5	19.6	18408	9 US-09-764-891-7441	Sequence 7441, App
19	104.5	19.6	18408	9 US-09-764-891-7442	Sequence 7442, App
20	100.5	18.9	88191	10 US-09-799-799-3	Sequence 3, Appl
21	100	18.8	3424	9 US-09-764-891-8887	Sequence 8887, App
22	95	17.9	752	9 US-10-091-504-1989	Sequence 1989, App
23	95	17.9	752	10 US-09-764-869-1989	Sequence 8874, App
24	95	17.9	2138	9 US-09-764-891-8874	Sequence 24, Appl
25	93.5	17.6	36303	9 US-10-152-724A-24	Sequence 1, Appl
26	93	17.5	28690	9 US-10-010-802-1	Sequence 9685, App
27	93	17.5	31813	9 US-09-764-891-9685	Sequence 2385, App
28	92	17.3	16886	10 US-09-764-891-2385	Sequence 3949, App
29	92	17.3	76798	10 US-09-880-107-3949	Sequence 3, Appl
30	92	17.3	203654	10 US-09-820-905-3	Sequence 1963, App
31	91.5	17.2	32169	9 US-10-092-154-1963	Sequence 1963, App
32	91.5	17.2	32169	10 US-09-764-847-1963	Sequence 1, Appl
33	91.5	17.2	465237	10 US-09-933-267A-1	Sequence 8888, App
34	91	17.1	3425	9 US-09-764-891-8888	Sequence 242, App
35	91	17.1	10236	9 US-10-091-438-282	Sequence 282, App
36	91	17.1	10236	9 US-10-091-438-282	Sequence 878, App
37	91	17.1	10236	10 US-09-764-853-878	Sequence 1, Appl
38	91	17.1	52216	10 US-09-747-810-1	Sequence 1429, App
39	90.5	17.0	7032	9 US-10-092-154-1429	Sequence 1429, App
40	90.5	17.0	7032	10 US-09-764-847-1429	Sequence 1735, App
41	90.5	17.0	7537	9 US-10-091-504-1735	Sequence 1735, App
42	90.5	17.0	7537	10 US-09-764-869-1735	Sequence 1735, App
43	90	16.9	63000	10 US-09-780-172-18	Sequence 18, Appl
44	89.5	16.8	601	10 US-09-820-003A-22	Sequence 22, Appl
45	89.5	16.8	601	10 US-09-820-003A-23	Sequence 23, Appl

## ALIGNMENTS

RESULT 1  
US-09-854-133-442  
; Sequence 442, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Serlist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 442  
; LENGTH: 337  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-854-133-442

Alignment Scores:  
Pred. No.: 2.77e-62 Length: 337  
Score: 532.00 Matches: 97  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-854-133-586 (1-97) x US-09-854-133-442 (1-337)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 |||||  
 Db 5 GAGGTTGAAGTGACGACAGATCATGCCAGCTGGGTGACAGTGTCTCTCAACA 64  
 |||||

QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysPheGlnAlaAsnCysGly 40  
 |||||  
 Db 65 GAATTAAGGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 124  
 |||||

QY 41 IleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGln 60  
 |||||  
 Db 125 ATAGATTTATCATATCTCGATTTTGGATTCTTTTGGATTCTCTCATCTGATTCAG 184  
 |||||

QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 |||||  
 Db 185 GAAAGCTGTGTGTCCACCATCTCCAAGAGAGGTTACCTGCAGGGAATGTTAACGGGA 244  
 |||||

QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
 |||||  
 Db 245 GGCTGCTTCCCTGGGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 295  
 |||||

## RESULT 2

US-09-738-973-442  
 ; Sequence 442, Application US/09738973  
 ; Patent No. US20020110563A1

## GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indrias, Carol Yoseph  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Elliot, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14  
 ; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 442  
 ; LENGTH: 337  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens

## US-09-738-973-442

Alignment Scores:  
 Pred. No.: 2,77e-62 Length: 337  
 Score: 532.00 Matches: 97  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-09-854-133-586 (1-97) x US-09-738-973-442 (1-337)

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 |||||  
 Db 5 GAGGTTGAAGTGACGACAGATCATGCCAGCTGGGTGACAGTGTCTCTCAACA 64  
 |||||

QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysPheGlnAlaAsnCysGly 40  
 |||||  
 Db 65 GAATTAAGGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 124  
 |||||

QY 41 IleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGln 60  
 |||||  
 Db 125 ATAGATTTATCATATCTCGATTTTGGATTCTTTTGGATTCTCTCATCTGATTCAG 184  
 |||||

QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 |||||  
 Db 185 GAAAGCTGTGTGTCCACCATCTCCAAGAGAGGTTACCTGCAGGGAATGTTAACGGGA 244  
 |||||

QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
 |||||  
 Db 245 GGCTGCTTCCCTGGGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 295  
 |||||

## RESULT 3

US-09-854-133-441  
 ; Sequence 441, Application US/09854133  
 ; Publication No. US20020183499A1

## GENERAL INFORMATION:

; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Secrist, Heather  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C10  
 ; CURRENT APPLICATION NUMBER: US/09/854,133  
 ; CURRENT FILING DATE: 2001-05-11  
 ; NUMBER OF SEQ ID NOS: 735  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 441  
 ; LENGTH: 5981  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens

## US-09-854-133-441

Alignment Scores:  
 Pred. No.: 6.32e-60 Length: 5981  
 Score: 527.00 Matches: 96  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.06% Indels: 0  
 DB: 9 Gaps: 0

US-09-854-133-586 (1-97) x US-09-854-133-441 (1-5981)

QY 2 ValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThrGlu 21  
 |||||  
 Db 3 GTTGAAGTGACGACAGATCATGCCAGCTGGGTGACAGTGTCTCTCTCAACAGAA 62  
 |||||

QY 22 LeuArgLysLysGluArgLysLysLysLysLysLysPheGlnAlaAsnCysGlyIle 41  
 |||||  
 Db 63 TTAAGGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATA 122  
 |||||

QY 42 AspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisHisTrpIleGlnGlu 61  
 |||||  
 Db 123 GATTTTATCATATCTCGATTTTGGATTCTTTTGGATTCTCTCATCTGATTCAGAA 182  
 |||||

QY 62 SerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGlyGly 81  
 |||||  
 Db 183 AGCCTGTGTGTCCACCATCTCCAAGAGAGGTTACCTGCAGGGAATGTTAACGGGAGC 242  
 |||||

QY 82 CysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
 |||||  
 Db 243 TGCCTTCCCTGGGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 290  
 |||||

## RESULT 4

US-09-738-973-441  
 ; Sequence 441, Application US/09738973  
 ; Patent No. US20020110563A1

## GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.

US-09-854-133-440	Alignment Scores:	1.24e-58	Length: 2239
US-09-854-133-440	Score: 513.00	Matches: 97	
	Percent Similarity: 84.35%	Conservative: 0	
	Best Local Similarity: 84.35%	Mismatches: 0	
	Query Match: 96.43%	Indels: 18	
	DB: 9	Gaps: 1	
US-09-854-133-586 (1-97) x US-09-854-133-440 (1-2239)			
QY 1 GluValGluValSerArgaspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20			
DB 2 GAGTTGAAGTGACGAGATCATGCCAGCTGGGTGACAGTCTCTCTCAACA 61			
QY 21 GluLeuArgLysLysGluArgLysLysLysLysLysLysLysLysLysLysLys 40			
DB 62 GAATTAAGGAAAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121			
QY 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHistrpIleGln 60			
DB 122 ATAGATTTATCATATCTCGATTTTGGATTTCTTTTGTCTCATCATCGATTTCAG 181			
QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80			
DB 182 GAAAGCGTGTGTCTCCACCATCTCCAAAGGAGGTACCTGCAGGAGAAATGTTACGGGA 241			
QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg 93			
DB 242 GGCTGCTTCCCTGGGCAACAGAGAGCCACCTGGGAGAGAGAGAGAGAGAGAGAG 301			
QY 94 -----ArgLysCysSer 97			
DB 302 CCTTTTCAGGAAG 346			
RESULT 6			
US-09-738-973-440			
; Sequence 440, Application US/09738973			
; Patent No. US20020110563A1			
; GENERAL INFORMATION:			
; APPLICANT: Reed, Steven G.			
; APPLICANT: Henderson, Robert A.			
; APPLICANT: Lodes, Michael J.			
; APPLICANT: Fling, Steven P.			
; APPLICANT: Mohamath, Raodoh			
; APPLICANT: Algate, Paul A.			
; APPLICANT: Secrist, Heather			
; APPLICANT: Indrias, Carol Yoseph			
; APPLICANT: Benson, Darin R.			
; APPLICANT: Elliott, Mark			
; APPLICANT: Mannion, Jane			
; APPLICANT: Kalos, Michael D.			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR			
; FILE REFERENCE: 210121.475C9			
; CURRENT APPLICATION NUMBER: US/09/738,973			
; CURRENT FILING DATE: 2000-12-14			
; NUMBER OF SEQ ID NOS: 587			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 440			
; LENGTH: 2239			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-738-973-440			
Alignment Scores:			
Pred. No.: 1.24e-58	Length: 2239		
Score: 513.00	Matches: 97		
Percent Similarity: 84.35%	Conservative: 0		
Best Local Similarity: 84.35%	Mismatches: 0		
Query Match: 96.43%	Indels: 18		
DB: 9	Gaps: 1		

```
US-09-854-133-586 (1-97) x US-09-738-973-440 (1-2239)
QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20
Db 2 GAGGTGAAGTGCAGCAGATCATGCCAGCCTGGGTGACAGTGAGACTGCTCTCAACA 61
QY 21 GluLeuArgLysGluArgLysLysArgGluArgLysPheGlnAlaAsnCysGly 40
Db 62 GAATTAAAGGAAAGAAAGAAAGAAAGAGAGAGAGAGAAATTCAGGCAATTTGGC 121
QY 41 IleAspPheIlePheThrPheThrPheThrPheThrPheThrPheThrPheThr 60
Db 122 ATAGATTATCATATTCGATTTTGGATCTTTTCTCATCATCATCATCATCATCAT 181
QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80
Db 182 GAAAGCCCTGTTGTCCACCCTCCAAAGAGAGTTTACCTGAGGAAATGTTAAAGGGA 241
QY 81 GlyCysLeuProThrAlaThrArgSerHisLeuGlyArg-----93
Db 242 GGCTGCTTCCCTGGGCAACAGAGGACCTGGGCGAGGAGCCCTTTTCAGGAAGAGAGC 301
QY 94 -----ArgLysCysSer 97
Db 302 CCTTTTCAGGAAGAGAGCCCTTTTCAGGAAGAGAGAGAGAGAGAGAGAGAGC 346
RESULT 7
US-10-046-935-1307
; Sequence 1307, Application US/10046935
; Patent No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046.935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1307
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9, 19, 461, 497, 500, 502
; OTHER INFORMATION: n = A,T,C or G
US-10-046-935-1307
Alignment Scores:
Pred. No.: 2,04e-20 Length: 572
Score: 221.00 Matches: 47
Percent Similarity: 70.00% Conservative: 9
Best Local Similarity: 58.75% Mismatches: 18
Query Match: 41.54% Indels: 6
DB: Gaps: 2
US-09-854-133-586 (1-97) x US-10-046-935-1307 (1-572)
QY 19 GlnThrGluLeuArgLysLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37
Db 142 CAGCTGAAAGCAGAGAGACATCATCATCATCATCATCATCATCATCATCATCATCAT 201
QY 38 AsnCysGlyIleAspPheIlePheThrPheThrPheThrPheThrPheThrPheThr 57
Db 202 TTTGTTTTCTTCCCTCTGTTTTATTTTCCCGCTGTGCTTACTATGTCA-----255
QY 58 TrpIleGlnGluSerLeuLeuCysProSerProLysGluValThrCysArgGluMet 77
Db 256 -----GAAAGCCTGTTGTCTCCACCATCTCCAAAGGAGGTTACCTGCAGGAAATG 306
QY 78 LeuThrGlyCysLeuProThrAlaThrArgSerHisLeuGlyArgLysCysSer 97
Db 307 TTAACGGGAGGCTGCTTCCCTGGGCAACAAAGAGGACCATCTGGGCGAGAGAGAGAG 366
RESULT 9
US-10-146-502-1307
; Sequence 1307, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146.502
; CURRENT FILING DATE: 2002-05-14
```

```

; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1307
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9, 19, 461, 497, 500, 502
; OTHER INFORMATION: n = A,T,C or G
US-10-146-502-1307

```

```

Alignment Scores:
Pred. No.: 2,04e-20 Length: 572
Score: 221.00 Matches: 47
Percent Similarity: 70.00% Conservative: 9
Best Local Similarity: 58.75% Mismatches: 18
Query Match: 41.54% Indels: 6
DB: Gaps: 2

```

US-09-854-133-586 (1-97) x US-10-146-502-1307 (1-572)

```

QY 19 GlnThrGluLeuArgLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37
||||:||||| ||| :||| |||:|||||
Db 142 CAGTCTGAAGCAGAGGAGACATCGATCAGTAACCAAGAGACACCAAGTTGAAAGT 201
||||:||||| ||| :||| |||:|||||
QY 38 AsnCysGlyLeuAspPheIlePheTrpIlePheTrpIleLeuPheSerHisHis 57
||||:||||| ||| :||| |||:|||||
Db 202 TTTGTTTCTTCTTCTCTGTTTATTTTCCCGCGTGTGCTTACTATGTGTC 255
||||:||||| ||| :||| |||:|||||
QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77
||||:||||| ||| :||| |||:|||||
Db 256 -----GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACTCTCAGGGAATG 306
||||:||||| ||| :||| |||:|||||
QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97
||||:||||| ||| :||| |||:|||||
Db 307 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGGAGGACCTGGGCGAGGAAAGTGCAGC 366
||||:||||| ||| :||| |||:|||||

```

```

RESULT 10
US-10-163-866-30
; Sequence 30, Application US/10163866
; Publication No. US20030027188A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: SLCT's AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-080C
; CURRENT APPLICATION NUMBER: US/10/163,866
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/338,733
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/357,600
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-163-866-30

```

```

Alignment Scores:
Pred. No.: 1.01e-19 Length: 1861
Score: 221.00 Matches: 47
Percent Similarity: 70.00% Conservative: 9
Best Local Similarity: 58.75% Mismatches: 18
Query Match: 41.54% Indels: 6
DB: Gaps: 2

```

US-10-163-866-30

```

Alignment Scores:
Pred. No.: 1.01e-19 Length: 1861
Score: 221.00 Matches: 47
Percent Similarity: 70.00% Conservative: 9
Best Local Similarity: 58.75% Mismatches: 18
Query Match: 41.54% Indels: 6
DB: Gaps: 2

```

```

US-09-854-133-586 (1-97) x US-10-163-866-30 (1-1861)
QY 19 GlnThrGluLeuArgLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37
||||:||||| ||| :||| |||:|||||
Db 129 CAGTCTGAAGCAGAGGAGACATCGATCAGTAACCAAGAGACACCAAGTTGAAAGT 188
||||:||||| ||| :||| |||:|||||
QY 38 AsnCysGlyLeuAspPheIlePheTrpIlePheTrpIleLeuPheSerHisHis 57
||||:||||| ||| :||| |||:|||||
Db 189 TTTGTTTCTTCTTCTCTGTTTATTTTCCCGCGTGTGCTTACTATGTGTC 242
||||:||||| ||| :||| |||:|||||
QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77
||||:||||| ||| :||| |||:|||||
Db 243 -----GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACTCTCAGGGAATG 293
||||:||||| ||| :||| |||:|||||
QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97
||||:||||| ||| :||| |||:|||||
Db 294 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGGAGGACCTGGGCGAGGAAAGTGCAGC 353
||||:||||| ||| :||| |||:|||||

```

```

RESULT 11
US-10-163-866-34
; Sequence 34, Application US/10163866
; Publication No. US20030027188A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: SLCT's AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-080C
; CURRENT APPLICATION NUMBER: US/10/163,866
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/338,733
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/357,600
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-163-866-34

```

```

Alignment Scores:
Pred. No.: 1.11e-19 Length: 2000
Score: 221.00 Matches: 47
Percent Similarity: 70.00% Conservative: 9
Best Local Similarity: 58.75% Mismatches: 18
Query Match: 41.54% Indels: 6
DB: Gaps: 2

```

US-09-854-133-586 (1-97) x US-10-163-866-34 (1-2000)

```

QY 19 GlnThrGluLeuArgLysGluArgLys---LysLysArgGluArgLysPheGlnAla 37
||||:||||| ||| :||| |||:|||||
Db 29 CAGTCTGAAGCAGAGGAGACATCGATCAGTAACCAAGAGGAGGTTACTCTCAGGGAATG 88
||||:||||| ||| :||| |||:|||||
QY 38 AsnCysGlyLeuAspPheIlePheTrpIlePheTrpIleLeuPheSerHisHis 57
||||:||||| ||| :||| |||:|||||
Db 89 TTTGTTTCTTCTTCTCTGTTTATTTTCCCGCGTGTGCTTACTATGTGTC 142
||||:||||| ||| :||| |||:|||||
QY 58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77
||||:||||| ||| :||| |||:|||||
Db 143 -----GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACTCTCAGGGAATG 193
||||:||||| ||| :||| |||:|||||
QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97
||||:||||| ||| :||| |||:|||||
Db 194 TTAACGGGAGGCTGCCTTCCCTGGGCAACAAGGAGGACCTGGGCGAGGAAAGTGCAGC 253
||||:||||| ||| :||| |||:|||||

```

```
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-163-866-32

Alignment Scores:
Pred. No.:          4 53e-19              Length:      520
Score:             210.50                Matches:       38
Percent Similarity: 95.00%               Conservative:   0
Best local Similarity: 95.00%            Mismatches:    1
Query Match:       39.57%                 Indels:        1
DB:                                                         Gaps:         1

US-09-854-133-586 (1-97) x US-10-163-866-32 (1-520)

QY  58 TrpIleGlnGluSerLeuLeuCysProProSerProLysGluValThrCysArgGluMet 77
    |||                               |||||
Db  2 TGG----TCAGAAAGCGCTGTGTGTCCACCATCTCCAAAGGAGGTTACCTGCAGGGAATG 58

QY  78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgLysCysSer 97
    |||                               |||||
Db  59 TTACGGGAGGCTGCCTTCCCCTGGGCCAACAGGAGCACCTGGGCAGGAGAAAGTGACG 118

RESULT 14
US-10-163-866-53
; Sequence 53, Application US/10163866
; Publication No. US20030027188A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: SLIC/5 AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
PRTN
```

[illegible]



RESULT 15  
US-10-163-866-52  
; Sequence 52, Application US/10163866  
; Publication No. US20030027188A1  
; GENERAL INFORMATION:  
; APPLICANT: EXELIXIS, INC.  
; TITLE OF INVENTION: SLCTs AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
; FILE REFERENCE: EX02-080C  
; CURRENT APPLICATION NUMBER: US/10/163,866  
; CURRENT FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: US 60/296,076  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/328,605  
; PRIOR FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/338,733  
; PRIOR FILING DATE: 2001-10-22  
; PRIOR APPLICATION NUMBER: US 60/357,253  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: US 60/357,600  
; PRIOR FILING DATE: 2002-02-15  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 52  
; LENGTH: 1528  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-163-866-52

Alignment Scores:  
Pred. No.: 1.95e-18 Length: 1528  
Score: 210.50 Matches: 38  
Percent Similarity: 95.00% Conservative: 0  
Best Local Similarity: 95.00% Mismatches: 1  
Query Match: 39.57% Indels: 1  
DB: 9 Gaps: 1

US-09-854-133-586 (1-97) x US-10-163-866-52 (1-1528)  
QY 58 TrpIleGlnGluSerLeuLeuCysProSerProLysGluValThrCysArgGluMet 77  
Db 24 TGG---TCAGAAAGCCTGTTGTGCCACCATCTCCAAAGGAGGTTACCTGCAGGGAATG 80  
QY 78 LeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 81 TTAACGGGAGGCTGCCTTCCTGGCAACAGAGGCCACCTGGGGCAGGAGAAAGTGCAGC 140

Search completed: May 11, 2003, 16:28:53  
Job time : 120.027 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 14:07:50 ; Search time 1470.45 Seconds  
(without alignments)  
1068.354 Million cell updates/sec

Title: US-09-854-133-586

Perfect score: 532

Sequence: 1 EWEVSRDHASIGDSETLSTQ.....ITGGCLPWATRSHLGRKCS' 97

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=cgn2.1/USPTO.spool/US09854133/runat\_05052003\_173956\_349/app\_query.fasta\_1.462  
-DB=EST -OFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCI=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09854133.cgn\_1\_1\_2013@runat\_05052003\_173956\_349 -NCPUP=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_man.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rod.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	221	41.5	936	12	BG284503
2	221	41.5	1072	12	BG388107
3	205	38.5	557	17	AQ002318
4	197.5	37.1	123	9	AI313891
5	194	36.5	910	12	BG326527
6	144.5	27.2	633	10	BG630988
7	144.5	27.2	652	10	BG612892
8	137.5	25.8	680	10	BG630675
9	131	24.6	524	12	BF286052
10	109.5	20.6	492	17	BH859880
11	109	20.5	669	17	AG037205
12	108.5	20.4	649	17	AQ629890
13	107.5	20.2	599	17	AQ541929
14	104.5	19.6	430	17	AQ414005
15	102	19.2	919	14	BQ277344
16	101.5	19.1	492	17	AQ008942
17	100.5	18.9	777	17	AF102045
18	100.5	18.9	832	13	BM011880
19	100	18.8	323	9	AA663028
20	99.5	18.7	404	17	AQ596495
21	99.5	18.7	531	17	AQ822937
22	99	18.6	688	17	AG087340
23	98.5	18.5	708	17	AG119431
24	98	18.4	828	12	BG281385
25	97.5	18.3	677	17	AG147658
26	96.5	18.1	562	17	AQ792364
27	96	18.0	298	12	BF866615
28	96	18.0	673	9	AL696209
29	95.5	18.0	518	10	AW849972
30	95.5	18.0	586	10	AV764383
31	95	17.9	350	12	BF095483
32	95	17.9	501	17	B99029
33	93.5	17.6	226	10	AW798344
34	93.5	17.6	917	12	BE733566
35	93	17.5	290	12	BF927188
36	93	17.5	599	17	AQ626336
37	93	17.5	649	17	AG038593
38	93	17.5	1112	13	BM450518
39	92.5	17.4	258	12	BF959970
40	92.5	17.4	716	12	BF676985
41	92	17.3	417	9	AA424117
42	91.5	17.2	502	9	AA458705
43	91	17.1	489	9	AA780515
44	91	17.1	1006	12	BG707662
45	90.5	17.0	385	17	AQ029438

# ALIGNMENTS

RESULT 1  
BG284503  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

BG284503  
602408645F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4537810 5', mRNA sequence.  
BG284503  
BG284503.1 GI:13035516  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 936)  
NIH-MGC <http://mgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)

936 bp mRNA linear EST 21-FEB-2001





Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh,  
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura,  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

Riken Gene Bank Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAAGATCCCAAGAGCTCTTTTGTGTTTTTN 3'], cDNA was prepared by using trisaccharose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGATTAAATTAATTCCTCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk existing.

strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTAAATTAAATAATCCGCCCGCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda  
FLC I." 144 a 154 c 175 g 159 t 1 others

```

es:
0.000657
144.50
ity: 51.06%
arity: 40.43%
27.16%
10
10
Gaps: 3
Length: 633
Matches: 38
Conservative: 10
Mismatch: 33
Indels: 13
86 (1-97) x BB630988 (1-633)
YAspSerGluThrLeuSerGlnThrGluLeuArgLysLysGluArgLysLys 30
|||||:
TGATAGCAAAGGGGAAGTCACGACCGACGAGTCATGCTACTTCTTAGAGAAC 229
|||||:

```

...GATCAGTCAGTCTACTTCTTAGAGAAC 229

## FEATURES





100

Best Local Similarity: 47.95%

Best Local Similarity: 47.95% Mismatches: 11

Query Match:	20.49%	Indels:	18
DB:	17	Gaps:	4
US-09-854-133-586 (1-97) x AG037205 (1-669)			
QY	1	GluValGluValSerArgAspHisAla	-----SerLeuGlyasp--s 14
Db	227	GAAGTTTCAGTGCCGAGATCATGCCACCATCTCCAGCACTCCAGCTCGGGTGACAGA	286
QY	14	erGluThrLeuSerGlnThrGluLeuArgLysLysGluArgLysLysArgGluArgL 34	
Db	287	GTGAGACTCTGTCTTTACCAAAA---AAAAAAGAAAAGAAAAGAAAAGAAAAGAA	343
QY	34	ysPheGlnAlaAsnCysGlyLeuAspPheIlePheTrpIlePheTrpLeuLeu 54	
Db	344	AA-----TGTGTGATA-----TATATACACCATGAATACTAT	376
QY	54	heserHisHsTrpIleGlnGluSerLeuLeuCys 65	
Db	377	TCAGCCATAAAAGGAACAAATAATGTTATTTGT	411
RESULT 12			
LOCUS AQ629890			
DEFINITION RPci-11-483N13.TV RPci-11 Homo sapiens genomic clone RPci-11-483N13			
ACCESSION AQ629890			
VERSION AQ629890.1 GI:5092525			
KEYWORDS GSS			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE 1 (bases 1 to 649)			
AUTHORS Zhao,S., Adams,W.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.			
TITLE Use of BAC End Sequences from Library RPci-11 for Sequence-Ready Map Building			
JOURNAL Unpublished (1997)			
COMMENT Other GSSs: RPci-11-483N13.TJ			
Contact: Shaying Zhao, William Nierman, Mark Adams			
Department of Eukaryotic Genomics			
The Institute for Genomic Research			
9712 Medical Center Dr., Rockville, MD 20850			
Tel: 301 838 0200			
Fax: 301 838 0208			
Email: hbe@tigr.org			
Clones are derived from the human BAC library RPci-11. For BAC library availability, please contact Pieter de Jong (pieter@ejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.			
Seq primer: n7			
Class: BAC ends.			
FEATURES			
Location/Qualifiers			
1..649			
/organism="Homo sapiens"			
/db_xref="GDB:7685412"			
/db_xref="taxon:9606"			
/clone="RPci-11-483N13"			
/clone_lib="RPci-11"			
/sex="Male"			
/cell_type="Lymphocytes"			
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPci11 Human Male BAC Library"			
BASE COUNT	202 a	116 c	143 g 188 t
ORIGIN			
Alignment Scores:			
Pred. No.:	1.43	Length:	649
Score:	108.50	Matches:	38
Percent Similarity:	42.24%	Conservative:	11



prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:  
 5'-AAGCAGTGGTATCAAGCAGAGTGGCCATACGGCCGG-3' and  
 5'-ATTCTAGAGCGAGCGCCGACATG-DT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART Kit and size-selected to contain the 1-2 kb size fraction (other fractions present in NIH\_MGC\_126 and NIH\_MGC\_128). Library created in the laboratory of T. Usdin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH\_MGC Library."

BASE COUNT 302 a 212 c 215 g 190 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	7.48	Length:	919
Score:	102.00	Matches:	38
Percent Similarity:	45.61%	Conservative:	14
Best Local Similarity:	33.33%	Mismatches:	17
Query Match:	19.17%	Indels:	45
DB:	14	Gaps:	6

US-09-854-133-586 (1-97) x BQ277344 (1-919)

```

QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAsp--SerGluThrLeuSerGlnT 20
Db 17 GAGGTTGCAGGAGCCCAAGATCGTGCCAGCCTGGTGACAGAGTGAGACTCTGTCTCAA 76
QY 20 hrGluLeuArgLysLysGluArgLysLysLysLysLysLysLysLysLysLysLysLys 40
Db 77 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120
QY 40 LysLeuAspPheIleIlePheTrpIlePheTrpIleLeu-LeuPheSerHis----- 56
Db 121 -----AGCTGGGTATGGTGGCTTATGCTGTAGTCGACGACTACTCAA 163
QY 57 -----HisTrpIleGlnGluSerLeuCysProProSerProLysGluValThrCys 74
Db 164 AGGCTGAGGTGG-----GAGGATTGCTTGTCCCC----- 193
QY 75 ArgGluMetLeuThrGlyGly-----CysLeuPro 84
Db 194 -----GGAGTTGAGCTACAGTGAGCCTTGATGTGCTCACTGCACTCCA 238
QY 85 ---TrpAlaThrArgSerHisLeuGlyArgArgLys 95
Db 239 GCCTGGGCAACAGGTAAACTCTGTCTCAAAAAA 274

```

Search completed: May 11, 2003, 15:28:01  
 Job time : 1476.45 secs

Tue May 13 12:12:49 2003

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: May 11, 2003, 18:33:52 ; Search time 1548 Seconds  
(without alignments)  
1823.625 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGSETLSQT.....LTGGCLPWATSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 0.0 , Xgapext 0.5  
Ygapop 0.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0-

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 3913070

Minimum DB seq length: 0  
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV-xlh  
-Q=/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_174426\_1210/app\_query.fasta\_1.263  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=5000 -NCPU=6 -ICPU=3  
-USER=US09854133 @CNI\_1.1687@runat\_05052003\_174426\_1210 -DEV\_TIMEOUT=120  
-NO\_XLPAY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=7  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=0 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_lm.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rtd.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgtgo\_hum.\*  
40: em\_hgtgo\_mus.\*  
41: em\_hgtgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	532	100.0	337	6	AX321911	Sequence
2	523	98.3	239	6	AX321909	Sequence
3	339.5	63.8	1683	9	AB042201	Homo sapi
4	339.5	63.8	1861	9	AB026891	Homo sapi
5	339.5	63.8	1874	9	AF200708	Homo sapi
6	339.5	63.8	2482	9	AF252872	Homo sapi
7	339	63.7	575	6	AX341060	Sequence
c 8	327.5	61.6	596	6	AX351341	Sequence
c 9	307.5	57.8	2155	9	BC012087	Homo sapi
c 10	297	55.8	927	10	AY065541	Mus muscu
c 11	297	55.8	4845	9	H0MEDHB17	Human esra
c 12	296.5	55.7	2000	9	AB040875	Homo sapi
c 13	294	55.3	1666	9	AC104107	Homo sapi
c 14	293.5	55.2	1671	10	BC022180	Mus muscu
c 15	293	55.1	4068	10	AF230072	Rattus no
c 16	292.5	55.0	3995	6	AX375015	Sequence
c 17	292.5	55.0	3995	6	AX375015	Sequence
c 18	291	54.7	4623	10	BC034547	Homo sapien
c 19	290	54.5	1600	8	AF019376	Brassica
c 20	289.5	54.4	3994	9	AL750302	Human DNA
c 21	289	54.3	2382	9	AK023017	Homo sapi
c 22	288	54.1	1269	3	AF160879	Drosophil
c 23	288	54.1	2385	9	HSA300588	Homo sapi
c 24	287.5	54.0	2427	6	AX001017	Sequence
c 25	287.5	54.0	2427	6	BD005487	G-coupled
c 26	287	53.9	900	9	HSGCKR1N1	Y09593 H.sapiens G
c 27	286.5	53.9	4314	9	HSM804205	AL832894 Homo sapi
c 28	286	53.8	4363	9	HSM804473	AL833162 Homo sapi
c 29	285.5	53.7	1906	4	AF339909	Capra hir
c 30	285.5	53.7	3848	9	HSL81672	AL81672 Homo sapien
c 31	285	53.6	3608	10	BC028880	BC028880 Mus muscu
c 32	284.5	53.5	938	10	MMU04333	U04333 Mus musculu
c 33	284.5	53.5	1114	9	HUMF13A01	M21986 Human coagu
c 34	284.5	53.5	1811	9	AK025185	AK025185 Homo sapi
c 35	284.5	53.5	2008	9	AK023483	AK023483 Homo sapi
c 36	284.5	53.5	2727	9	AK001575	AK001575 Homo sapi
c 37	284.5	53.5	4313	9	HSM800952	AL117425 Homo sapi
c 38	284	53.4	1310	17	AF116717	AF116717 Homo sapi
c 39	284	53.4	3267	9	AK027045	AK027045 Homo sapi
c 40	284	53.4	3890	14	RNMRS1	BC012925 Homo sapi
c 41	284	53.4	3267	9	BC012925	J04357 Red clover
c 42	283.5	53.3	1517	8	AY038318	AY038318 Inocybe 1
c 43	283.5	53.2	2533	6	AR122536	AR122536 Sequence
c 44	283	53.2	2590	9	AB061850S1	AB061850 Homo sapi
c 45	283	53.2	2599	10	AF068707	AF068707 Mus muscu

ALIGNMENTS

RESULT 1

AX321911	AX321911	337 bp	DNA	linear	PAT 15-DEC-2001
LOCUS	Sequence 442 from Patent WO0172295.				
DEFINITION	AX321911				
ACCESSION	AX321911				
VERSION	AX321911.1	GI:17906521			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1				
AUTHORS	Reed, S.G., Lodes, M.J., Mohamath, R., Secrist, H., Benson, D.R., Indrias, C.Y., Henderson, R.A., Fling, S.P., Algate, P.A., Elliot, M., Mannion, J. and Kalos, M.D.				
TITLE	Compositions and methods for the therapy and diagnosis of lung cancer				
JOURNAL	Patent: WO 0172295-A 442 04-OCT-2001;				
FEATURES	CORIXA CORPORATION (US)				
source	1. .337				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	103 a 60 c 93 g 81 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	2.97e-06	Length:	337		
Score:	532.00	Matches:	97		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	6	Gaps:	0		
US-09-854-133-586 (1-97) x AX321911 (1-337)					
QY 1	GLUValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20				
Db	5 GAGTTGAAGTGAGCAGAGATCATGCCAGCTGGTGACAGTGACACTCTGTCTCAACA 64				
QY 21	GLULeuArgLysLysGluArgLysLysLysLysLysLysLysLysLysLysLysLysLys 40				
Db	65 GAATTAAAGAAAAAGAAAGAAAGAAAGAAAGAGAGAGAGAAATTCAGGCCAATTGGC 124				
QY 41	ILEAspPheIleIlePheIlePheIlePheIlePheIlePheIlePheIlePheIlePhe 60				
Db	125 ATAGATTTTATCATATCTGGATTTTGGATCTTTTGGATCTTTTCTCACTGGATTTCAG 184				
QY 61	GLULeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80				
Db	185 GAAAGCCTGTTGTGTCCACCATCTCCAAAGGAGGTTACCTGTCAGGAGAAATGTTAACGGGA 244				
QY 81	GLYCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97				
Db	245 GGCTGCCTTCCTGGGCACACAGGAGCCACCTGGGCAGGAGAGAGATGTCAGC 295				
RESULT 2					
LOCUS	AX321909	2239 bp	DNA	linear	PAT 15-DEC-2001
DEFINITION	Sequence 440 from Patent WO0172295.				
ACCESSION	AX321909				
VERSION	AX321909.1	GI:17906515			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1				
AUTHORS	Reed, S.G., Lodes, M.J., Mohamath, R., Secrist, H., Benson, D.R., Indrias, C.Y., Henderson, R.A., Fling, S.P., Algate, P.A., Elliot, M., Mannion, J. and Kalos, M.D.				
TITLE	Compositions and methods for the therapy and diagnosis of lung cancer				
JOURNAL	Patent: WO 0172295-A 440 04-OCT-2001;				



ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AF200708  
AF200708.1 GI:11493651

ORGANISM

Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1874)  
Conklin, D.S. and Beach, D.H.  
CCBR1, novel CD98 light chain implicated in redox control and  
calcium signaling  
Unpublished  
2 (bases 1 to 1874)  
Conklin, D.S. and Beach, D.H.  
Direct Submission  
Submitted (01-NOV-1999) Cold Spring Harbor Labs, 1 Bungtown Rd,  
Cold Spring Harbor, NY 11724, USA

FEATURES

source

1..1874

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_line="fibrosarcoma HT1080"

246..1751

/note="X-c transporter; CD98 light chain"

/product="calcium channel blocker resistance protein

CCBR1"

/protein\_id="AAG35592.1"

/db\_xref="GI:11493652"

/translation="MVRKPVVSTISKGYLQGVNNGRLPSLGNKEPPQKQVQLKRV  
TLKSGVSIITIGAGIFISPKGVLTQSGVMSLTITWVCGVLSLFGALSALYELGT  
PELAIKLITAVGITVVMVLSMSVSRIOIFLTPCKTALIIIVPGVMQLKQGT  
QNFKDFSGRDSITRLPLAFYGYMAYAGWYLFNFVEVENPEKTIPLAIGISMAI  
VTIGYVLTNAVYTTINAEELLSNNAVAVTFSERLGNFSLAVPIFVLSLSCFSGMNGG  
VFASRLFFVASREGLPEILSKIHVHKHTPLPAVILHPLTMILFSGDLSLLNPL  
SFARWLFGLAVAGLYLRYKCPDMHRPFKPLFIPALSFCTCLFWALSLSYDPFST  
GIGFVITLTGPAYLFIITWDRKPRFRIMSEKITRTLIQLIILEVVPEDKL"

BASE COUNT

490 a 407 c 422 g 555 t

Alignment Scores:

Pred. No.: 446 Length: 1874  
Score: 339.50 Matches: 62  
Percent Similarity: 57.25% Conservative: 13  
Best Local Similarity: 47.33% Mismatches: 7  
Query Match: 63.82% Indels: 49  
DB: 9 Gaps: 28

US-09-854-133-586 (1-97) x AF200708 (1-1874)

QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
Db 16 GAG---GAGGTGGAGAAATGAGACGATGATACACAGGTGTTCTGAGTAGTAATA 72  
QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
Db 73 ---GATCGCTGTGAAGAAACACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 126  
QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
Db 127 CGCTGAGAGACAGCTGTAACACAGGAGACATCGATCAGTAACACACAGACACACC 186  
QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49  
Db 187 AAA-----GTTGAAAGTTT---GTTTCTTCTCCCTCTGTTTATTT 225  
QY 50 Trp-----Ile---LeuLeuPheSerHisHsTrpIleGlnGluSerLeuLeuCysPro 66  
Db 226 TTCCCGCGGTGTCCTACTA-----TGG-----TCAGAAAGCTGTGTGTCTCA 270  
QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
Db 1874

Db 271

CCATCTCCAAAGAGAGTTACCTCAGGGAATGTTAACGGGAGGCTCCCTTCCTGGGCA 330

QY 87

ThrArgSerHisLeuGlyArgArgLysCysSer 97

Db 331

ACAAGGAGCCACTGGGCGAGGAGAAAGTGCAGC 363

RESULT 6

AF252872

LOCUS

AF252872

DEFINITION

AF252872

ACCESSION

AF252872

VERSION

AF252872.1

KEYWORDS

GI:13924719

SOURCE

ORGANISM

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2482)

Chancy, C.D., Kekuda, R., Wang, H., Huang, W., Prasad, P.D., Smith, S.B.

and Ganapathy, V.

Structure, Function and Regulation of Human Cystine/Glutamate

Transporter in Retinal Pigment Epithelial Cells

Unpublished

2 (bases 1 to 2482)

Wang, H., Prasad, P.D. and Ganapathy, V.

Direct Submission

Submitted (05-APR-2000) Biochemistry & Molecular Biology, Medical

College of Georgia, 1120 15th Street, Augusta, GA 30912, USA

FEATURES

source

1..2482

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/tissue\_type="placenta"

232..1737

/codon\_start=1

/product="cystine/glutamate transporter xct"

/protein\_id="AAK4911.1"

/db\_xref="GI:13924720"

/translation="MVRKPVVSTISKGYLQGVNNGRLPSLGNKEPPQKQVQLKRV

TLKSGVSIITIGAGIFISPKGVLTQSGVMSLTITWVCGVLSLFGALSALYELGT

PELAIKLITAVGITVVMVLSMSVSRIOIFLTPCKTALIIIVPGVMQLKQGT

QNFKDFSGRDSITRLPLAFYGYMAYAGWYLFNFVEVENPEKTIPLAIGISMAI

VTIGYVLTNAVYTTINAEELLSNNAVAVTFSERLGNFSLAVPIFVLSLSCFSGMNGG

VFASRLFFVASREGLPEILSKIHVHKHTPLPAVILHPLTMILFSGDLSLLNPL

SFARWLFGLAVAGLYLRYKCPDMHRPFKPLFIPALSFCTCLFWALSLSYDPFST

GIGFVITLTGPAYLFIITWDRKPRFRIMSEKITRTLIQLIILEVVPEDKL"

BASE COUNT

735 a 484 c 521 g 742 t

ORIGIN

Alignment Scores:

Pred. No.: 917 Length: 2482

Score: 339.50 Matches: 62

Percent Similarity: 57.25% Conservative: 13

Best Local Similarity: 47.33% Mismatches: 7

Query Match: 63.82% Indels: 49

DB: 9 Gaps: 28

US-09-854-133-586 (1-97) x AF252872 (1-2482)

QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11

Db 2 GAG---GAGGTGGAGAAATGAGACGATGATACACAGGTGTTCTGAGTAGTAATA 58

QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22

Db 59 ---GATCGCTGTGAAGAAACACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 112

QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33

Db 113 CGCTGAGAGACAGCTGTAACACAGGAGACATCGATCAGTAACACACAGACACACC 172

QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49



us-09-854-133-586.ige

Tue May 13 12:12:49 2003

```

173 AAA-----GTTGAAAGTTT---GTTTCTTCCCTCTGTTTATT 211
50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66
212 TTCCCCCGTGTGCCCTACTA-----TGG---TCAGAAAGCTGTTGTGTCCA 256
67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyCysLeuProTrpAla 86
257 CCATCTCCAAAGAGGTTTACCTCAGGGAATGTTAACGGGAGGCTGCCCTCCCTGGGCA 316
87 ThrArgSerHisLeuGlyArgArgLysCysSer 97
317 ACAAGGAGCCCTGGGAGGAGAAAGTGCGC 349

RESULT 7
AX341060 AX341060 575 bp DNA linear PAT 10-JAN-2002
LOCUS
DEFINITION Sequence 1307 from Patent WO0196388.
ACCESSION AX341060
VERSION AX341060.1 GI:18137042
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Jiang, Y., Harlocker, S.L. and Secrist, H.
Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL CORIXA CORPORATION (US)
PATENT: WO 0196388-A 1307 20-DEC-2001;
FEATURES
source
Location/Qualifiers
1..575
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 151 a 122 c 153 g 141 t 8 others
ORIGIN
Alignment Scores:
Pred. No.: 22.3 Length: 575
Score: 339.00 Matches: 62
Percent Similarity: 57.23% Conservative: 13
Best Local Similarity: 47.33% Mismatches: 6
Query Match: 63.72% Indels: 50
DB: 6 Gaps: 28

US-09-854-133-586 (1-97) x AX341060 (1-575)
QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11
Db 22 GAGTGGAG-----AATGAGAGCAGCATGATACACAGGTGTTCTGAGTAGTAATTA 75
QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22
Db 76 ---GATCCCTGTGAGGAAAGAACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 129
QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33
Db 130 CGCTGAGAGACAGTCTGAAAGCAGAGAGACATCATGATCAGTAACACCAAGAGACACC 189
QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49
Db 190 AAA-----GTTCAAGTTT---GTTTCTTCCCTCTGTTTATT 228
QY 50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66
Db 229 TTCCCCCGTGTGCCCTACTA-----TGG---TCAGAAAGCTGTTGTGTCCA 273
QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyCysLeuProTrpAla 86
Db 274 CCATCTCCAAAGAGGTTTACCTCAGGGAATGTTAACGGGAGGCTGCCCTCCCTGGGCA 333

```

```

87 ThrArgSerHisLeuGlyArgArgLysCysSer 97
334 ACAAGGAGCCCTGGGAGGAGAAAGTGCGC 366

RESULT 8
AX351341/c AX351341 596 bp DNA linear PAT 06-FEB-2002
LOCUS
DEFINITION Sequence 88 from Patent WO0196390.
ACCESSION AX351341
VERSION AX351341.1 GI:18616688
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Jiang, Y., Hepler, W.T., Clapper, J.D., Wang, A. and Secrist, H.
Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL CORIXA CORPORATION (US)
PATENT: WO 0196390-A 88 20-DEC-2001;
FEATURES
source
Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 160 a 151 c 126 g 158 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 57.8 Length: 596
Score: 327.50 Matches: 59
Percent Similarity: 57.48% Conservative: 14
Best Local Similarity: 46.46% Mismatches: 5
Query Match: 61.56% Indels: 49
DB: 6 Gaps: 27

US-09-854-133-586 (1-97) x AX351341 (1-596)
QY 6 Arg-----AspHis---Ala-----Ser-----LeuGlyAsp---Ser 14
Db 595 AAGAATTGAGAGCAGCATGATCATACACAGGTGTTCTGAGTAGTAATTA---GATCGGTGT 539
QY 15 Glu-----Thr---LeuSer---Gln-----ThrGluLeu---Arg-----Lys 24
Db 538 GAAGNAAAGACACACCTTTGAGTTTTCACCTGTGAACA---CTATAGCGGTGAGAGAGA 482
QY 25 Lys---Glu---Arg---Lys---LysLys-----ArgGluArgLysPheGlnAla 37
Db 481 CAGTCTGAAAGCAGAGGAGGAGACATCATGATCATACACCAAGAGACACCAA--- 431
QY 38 AsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---PheTrp-----Ile 51
Db 430 -----GTTGAAAGTTT---GTTTCTTCCCTCTGTTTATTTCCTCCCGGTGTG 383
QY 52 ---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysProSerProLys 70
Db 382 TCCCTACTA-----TGG---TCAGAAAGCTGTTGTGTCACCATCTCCAAG 338
QY 71 GluValThrCysArgGluMetLeuThrGlyCysLeuLeuProTrpAlaThrArgSerHis 90
Db 337 GAGGTACCTGCAGGGAATGTTAACGGGAGGCTGCCCTCCCTGGGCAACAGGAGCCAC 278
QY 91 LeuGlyArgArgLysCysSer 97
Db 277 CTGGCAGGAGAAAGTGCGC 257

RESULT 9
BC012087 2155 bp mRNA linear PRI 06-AUG-2001
LOCUS
DEFINITION Homo sapiens, Similar to solute carrier family 7, (cationic amino
acid transporter, y+ system) member 11, clone MGC:20026
IMAGE:4562994, mRNA, complete cds.
ACCESSION BC012087

```



VVOAIIISPRASGRRRLRPKYAMHILFPLFFWVNLGLISVNLHSTNTNRLNISHPR  
 GDSNCVFMPSOKIKWIFLPLDAVFGQAGGAGSYMVFLLYKHQHVLYLQNS  
 KILYRPPLELRAAQSVLLMLCFVFFVWTDVFSFPLSLSEYSLMNVNHEFTLGY  
 AIPSFVLIHRDGLLEQWETLKQCVSHLSQ"

BASE COUNT 227 a 197 c 191 g 312 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.77e+03 Length: 927  
 Score: 297.00 Matches: 65  
 Percent Similarity: 27.61% Conservative: 17  
 Best Local Similarity: 21.89% Mismatches: 9  
 Query Match: 55.83% Indels: 206  
 DB: 10 Gaps: 56

US-09-854-133-586 (1-97) x AY065541 (1-927)

Qy 1 GluVal---Glu-----Val-----Ser-----Arg-----AspHis--- 8  
 Db 919 CAGATAGATGAGAGACATGTTTCAATGTCCTCCACTGTTTCAGTCAGAGACCATCTC 860  
 Qy 9 -----Ala-----Ser-----LeuGlyAspSerGluThr----- 16  
 Db 859 TGTGAATCAGCACAAGGGCTGAATTCGCATACCAAGGG---TCAGAACTCATGAA 803  
 Qy 17 ---Leu---Ser---GlnThrGluLeuArgLysLys-----GluArgLysLysLys--- 30  
 Db 802 TATTTACCATCAACGAATACTCTCTA---AAGAGAGACTTAGAA---AAAAAGAAACA 749  
 Qy 31 -----ArgLysLysLys-----Phe----- 35  
 Db 748 CACAATCTCTCAGTAGA---AGAAACAAACAGAGCATCAGAGGAGACACATTTGAG 692  
 Qy 36 Gln-----Ala-----Asn-----Cys-----Gly----- 40  
 Db 691 CAGCTCAGCTCAGGGGAGTCTCTAGAGAGCTTGGAGTCTCTGAAGTAGAGACAT 632  
 Qy 41 IleAspPheIle-----Ile-----PheTrp---IlePheTrp----- 50  
 Db 631 GTTGATGGTGTGTGTAGAGAAGAAATACCATGTAGCAGCTGGCCCTCCCATGGCACCT 572  
 Qy 51 -----Ile-----Leu-----LeuPhe----- 54  
 Db 571 GAACACTGCATCCCTTAGGACCATCAGAGGAGAAATCCATTTATTTCTGACTTG 512  
 Qy 55 -----SerHisHis----- 57  
 Db 511 GTAACATAAATAACAATTCCTCATCACCCCTAAAGTGTGATATATTCAGTCTGTAT 452  
 Qy 58 ---Trp-----LysGlu-----Ile----- 59  
 Db 451 TTGTGGTAGAATGGATTAGTTCACACTTATTAACCATTTAGTACCCAAAGAAATGACA 392  
 Qy 60 -----GlnGlu-----Ser-----LeuLeu--- 64  
 Db 391 AGAATGGAAGAATGTGCCATGCATCTTGGCTGAGCCCTCCCTCCATCCAGATGCTCTGG 332  
 Qy 65 -----Cys----- 65  
 Db 331 GACTGATGATGGCGCTGGACACATGAGGAGACTGTGCTGGTCAGATGACAGGCCCC 272  
 Qy 66 ---ProProSerPro---LysGlu-----ValThrCys----- 74  
 Db 271 GGGCCACCTCTCCAGTAACAATGCTCTTACAGCCTA---TGTCATCCAGGAAGTTTC 215  
 Qy 75 -----Arg-----Glu---MetLeu---ThrGly---GlyCysLeu 83  
 Db 214 TCAAAACCAAAAGCTGCTACTCTCTTTCGAAATCTTTTGTAAAGAGATTAGGATGTTG 155  
 Qy 84 -----Pro-----TirPalanArgSer---His---Leu--- 91  
 Db 154 TAAAGGCCAAGTGGATAGAATGAGGTGTATGGTTTCT---TCTTAACACCTTCTAATA 98

Qy 92 ---GlyArg-----Arg---Lys-----Cys---Ser 97  
 Db 97 AACGGAAGAAATATCCACAGAAACAGAAATATCCCGAGGTGCCATCA 47  
 RESULT 11  
 HUMEDHB17/c HUMEDHB17 4845 bp DNA linear PRI 11-OCT-1996  
 LOCUS Human estradiol 17 beta-dehydrogenase gene, complete cds.  
 DEFINITION M27138 M27136  
 ACCESSION M27138.1 GI:181950  
 VERSION M27138.1  
 KEYWORDS Alu repeat; estradiol 17-beta-dehydrogenase.  
 SOURCE Human, cDNA to mRNA, and leukocyte DNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 4149)  
 AUTHORS The V.I., Labrie, C., Zhao, H.F., Couet, J., Lachance, Y., Simard, J.,  
 Leblanc, G., Cote, J., Berube, D., Gagne, R., and Labrie, F.  
 TITLE Characterization of cDNAs for human estradiol 17 beta-dehydrogenase  
 and assignment of the gene to chromosome 17: evidence of two mRNA  
 species with distinct 5'-termini in human placenta  
 JOURNAL Mol. Endocrinol. 3 (8), 1301-1309 (1989)  
 MEDLINE 89384667  
 PUBMED 2779584  
 REFERENCE 2 (bases 1 to 4845)  
 AUTHORS Luu-The, V., Labrie, C., Simard, J., Lachance, Y., Zhao, H.F., Couet, J.,  
 Leblanc, G., and Labrie, F.  
 TITLE Structure of two in tandem human 17 beta-hydroxysteroid  
 dehydrogenase genes  
 JOURNAL Mol. Endocrinol. 4 (2), 268-275 (1990)  
 MEDLINE 90231340  
 PUBMED 2330005  
 COMMENT [1] exons only.  
 Draft entry and computer-readable sequence for [1] kindly submitted  
 by F. Labrie, 15-AUG-1989.  
 FEATURES  
 source  
 1. 4845  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17q11-q12"  
 /map="17q11-q21"  
 repeat\_region  
 113..392  
 /note="Alu repeat copy A"  
 repeat\_region  
 416..696  
 /note="Alu repeat copy B"  
 repeat\_region  
 711..950  
 /note="Alu repeat copy C"  
 prim\_transcript  
 910..4149  
 /note="EDHB17 mRNA (alt.)"  
 prim\_transcript  
 1862..4149  
 /note="EDHB17 mRNA (alt.)"  
 CDS  
 join(1871..1967,2061..2228,2376..2555,2731..2824,  
 3341..3518,3605..3874)  
 /note="estradiol 17 beta-dehydrogenase"  
 /codon\_start=1  
 /protein\_id="AAB16941.1"  
 /db\_xref="GI:181951"  
 gene  
 1871..1967  
 /gene="EDH17B1"  
 exon  
 <1871..1967  
 /gene="EDH17B1"  
 /note="estradiol 17 beta-dehydrogenase; G00-119-860"  
 /number=1  
 intron  
 1968..2060  
 /note="EDHB17 intron A"  
 exon  
 2061..2228

REFERENCE	1 (sites)
AUTHORS	Kim, J. Y., Chairoungdua, A., Cha, S. H., Segawa, H., Matsuo, H., Kim, D. K., Endou, H. and Kanai, Y.
TITLE	Human cystine/glutamate exchanger: cDNA cloning and upregulation by oxidative stress in glioma cells
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 2000)
AUTHORS	Kanai, Y.

TITLE	Direct Submission	Submitted (01-APR-2000)	Yoshikatsu Kanai
JOURNAL			

Shinkawa, Department of Pharmacology and Toxicology; 6-20-2  
Shinkawa, Mitaka, Tokyo 181-8611, Japan  
(E-mail: ykanai@kyorin-u.ac.jp, Tel: +81-422-47-5511 (ex. 3453),  
Fax: +81-422-79-1321)

```

source
location/Qualifiers
1. 2000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hvcn"

```

```

gene
/clone_lib="brain cDNA library"
/dev_stage="adult"
1..2000
/gene="hxCt"

```

cds  
136..1620  
/gene="hxCt"  
/note="Na<sup>+</sup>-independent amino acid transporter  
transmembrane protein"

```
/codon_start=1  
/product="cystine/glutamate exchanger"  
/protein_id="BAB40574.1"  
/db_xref="GI:13516846"  
/translation="MWDVDEGMCYCACTCAGATLGGTGLDQK"
```

[illegible][illegible]

REGISTRATION	305 A	418 C	435 g	642 t
Alignment Scores:				
Seq. No.:	1	324+04	12444	2000

Core:	296.50	Length:	2000
Percent Similarity:	58.88%	Matches:	57
1st Local Similarity:	53.27%	Conservative:	6
Every Match:	55.73%	Mismatches:	10
		Indels:	34

9 Gaps: 19

[illegible]



Tue May 13 12:12:49 2003

Identification of a novel non-lysosomal sulfatase expressed in the floor plate, choroid plexus, and cartilage

Unpublished  
2 (bases 1 to 4068)  
Ohto, T. and Masu, M.  
Direct Submission  
Submitted (31-JAN-2000) Molecular Neurobiology, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8575, Japan

Location/Qualifiers  
1. .4068  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"

414..3026  
/codon\_start=1  
/product="sulfatase Fp"  
/protein\_id="AAL71906.1"

/db\_xref="GI:18461371"  
/translation="MKNSWALLLAVLGAELGGFCSTMRQFRGRVQQRKNIRPN  
IILVLDODVSLQVNMKTRKIMEHGATFNATVFTPMCCPSRSSMLTKYVHN  
HNYTNNCSPPSQALHEPRTFAVLTNTGYRTAFKYLNEYKSGSYTPPGWRWL  
GLKNSRFNYTCVRNGLEKHEGFDVAKDYTDLTITNESINYPKMSKRMYPHPRVMV  
ISHAHPGDSAPQPSKLYPNASOHTPSYVAPNMDKHWIYTGQPLVPIHMEFTN  
VLOKRLQTLMSVDSVERLYNMLVETGELGNTYLYTADHGVHIGQFLVKGLSMEY  
DFDIRVFFIRGPSIEGPSIVPOVLNIDLAPILDIAGLDTPSDVDGKSVLKLDLE  
KPNRFTNKAKIWRDTFLVERKFLKKEESKNIOQNHLPSYRVELCOQARY  
OTACEOPQNMOCIEDTSGKLRHKCPDLDLTVRQNRNLYSRGLODKKECHRE  
SGYEPSRSORRKHROFGRHQAAGDFRNLADNSAVGLPTTVRTHKCFILPNOT  
LPSPSAKHDEGHOGFGHQAAGDFRNLADNSAVGLPTTVRTHKCFILPNOT  
IHCERELYQARAWKHAKYIDKIEVLQDKIKLNRVGRHLKRRKKEKRRKRGECSL  
NKEGVGRKELSHLPFKKAEQVDSKLOLKEHRRKKEKRRKRGECSCGQSY  
PGLTCTFDHNNHQTAPFNNSFCACCTSSNNYTWCLRTVNEHLECFEATGFE  
YFDMTPOIQTNTVHTVHTVERGINOLHQLMELRSCOGIKOCNRPKSLDVTREGGN  
YDPHNGQLWDGEG"

BASE COUNT 1132 a 1002 c 1060 g 874 t

ORIGIN

Alignment Scores:

Pred. No.: 1.04e+05 Length: 4068

Score: 293.00 Matches: 68

Percent Similarity: 25.30% Conservativity: 15

Best Local Similarity: 20.73% Mismatches: 11

Query Match: 55.08% Indels: 234

DB: 10 Gaps: 62

US-09-854-133-586 (1-97) x AF230072 (1-4068)

QY 1 GluValGluVal Ser ArgAsp His 8

Db 2355 GAGATTGAGTTCTACAGGATAAAATTAAGAAATTAAGGAGAGTGAGGGACACCTAAAG 2414

QY 9 AlaSerLeuGly AspSer Glu 15

Db 2415 AAAAGGAAGCCTGAGGAGTGTAGTGTGTAAGCAGAGCTATTACACAAGAGAAAGGT 2474

QY 16 Thr Leu Ser Gln 19

Db 2475 GTCAACGACGAGGAGAGCTGAAGAGCCACCTTCACCCCTCAAGAGAGGCTGCGGCCAG 2534

QY 20 ThrGlu Leu ArgLysLysGluArg 27

Db 2535 --GAGGTGGACAGCAAACTGCAGCTGTTCAGAGGAGCATGCTCGGAGGAAAGAGGAGG 2591

QY 28 Lys LysArgGluArgLysPheGlnAlaAsnCys GlyIle 41

Db 2592 AAGGAGAAGAACGTCAGAGAAAGAGGAAA --GAGTSCAGCCTGCCTGGCTCACCTGC 2648

QY 42 Asp Phe Ile Ile PheTrp Ile Phe 49

Db 2649 TTCACCCACGACAAACCACTGCGAGACCGCCGCTTCTGGAACCTGGGATCTTCTGT 2708

QY 50 Trp Ile Ile Ile Ile 51

Db 2709 GCATGCAGCAGTCTTCAACAACATACTACTGTGTTCCTGCTACGGTCAACGAGAGGCAC 2768

QY 52 Leu Leu Phe 54

Db 2769 AATTTTCTCTCTCTGAGTTTGTCTACTGGCTTCTTGGAGTATTGTGACATGACACACAGAT 2828

QY 55 Ser His His 57

Db 2829 CATTACAGCTTACAAACACGGTCCACACGGTAGACGGGSCATCTTGAATCAGTACAC 2888

QY 58 TrpIleGln Glu SerLeu Leu Cys Pro Pro 67

Db 2889 --ATACAGCTGATGGAGCTCCGGAAGCTGCCAAGGTTATAAACAAGTGCACACCAAGCCC 2945

QY 68 Ser Pro Lys 70

Db 2946 AAGAGCCTTGACCTTGGAACTAAAGAGGAGGAACTATGACCCACACAGAGGACAGTTA 3005

QY 71 Glu ValThrCys Arg GluMet 77

Db 3006 TGGGATGGATGGGAAGGTTAGTTGGTCCAGTGTCTTCAGACACACCACTGGCAAGGCT 3065

QY 78 Leu Thr 79

Db 3066 GGAGGAGTTATCCGGTGCACACGACATCAAGAGGACAGATCTAACCTTAGACTGAGGCC 3125

QY 80 Gly GlyCys Leu 83

Db 3126 GGAGCCTGGACCAATTACCTGAGGAGTGTCCACAGAGCCTTTGCACTGCTGAACAGTCA 3185

QY 84 Pro Trp Ala Thr Arg SerHis 90

Db 3186 CCCTGATCAAAACCAAGTAATGGAGTCCAACTGCACCAAGCGGTGGCTCCCACTCAC 3245

QY 91 Leu Gly Arg Arg Lys 95

Db 3246 CTCTGCTGAGCTTGTGTCGCGCAGAGATGGCTTCTGCAAGTCTAGGTGAAGACCCAG 3305

QY 96 Cys Ser 97

Db 3306 TCATGTGTTGGTGGAAACGCTCC 3329

Search completed: May 11, 2003, 19:32:35  
Job time : 1554 secs





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 17:20:55 ; Search time 156 Seconds  
(without alignments)  
1400.281 Million cell updates/sec

Title: US-09-854-133-586

Perfect score: 532  
Sequence: 1 EWEVSRDHASLGDSLTSTQ.....LTGGCLPWATRSHLGRKCS 97

Scoring table:

BLOSUM62	Xgapop	0.0	Xgapext	0.5
	Xgapop	0.0	Xgapext	0.5
	Fgapop	6.0	Fgapext	7.0
	Delop	6.0	Delext	7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4313688

Minimum DB seq length: 0

Maximum DB seq length: 5000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q/cgn2.1/USPTO\_spool/US09854133/runat\_05052003\_174425\_1202/app\_query.fasta\_1.263  
-DB-N Geneseq 101002 -QFMT=fastap -SUFFIX=ing -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -NAXLEN=5000  
-USER=US09854133 -CGN\_1\_1\_200 -runat\_05052003\_174425\_1202 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=0 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	532	100.0	337	23	Human lung tumour-
2	523	98.3	2239	23	Human lung tumour-
3	339.5	63.8	800	20	Human gene express
4	339	63.7	575	24	Human colon tumour
5	327.5	61.6	596	24	Human colon cancer
6	304	57.1	2752	22	Human nervous syst
7	298.5	56.1	1687	22	Human ovarian and
8	298.5	56.1	1687	22	Human reproductive
9	298.5	56.1	1688	22	Human ovarian and
10	298.5	56.1	1888	22	Human endometrial
11	292.5	55.0	3995	24	Gene DUG4 differen
12	292.5	55.0	3995	24	Human pancreatic c
13	292	54.9	2197	22	Human digestive sy
14	292	54.9	2197	22	Human nervous syst
15	291	54.7	2751	22	Human reproductive
16	289.5	54.4	2592	23	Drosophila melanog
17	289	54.3	2382	22	Human cDNA sequenc
18	287.5	54.0	2427	20	Human P2Y11 recept
19	286	53.8	2010	24	Human DNA replicat
20	285.5	53.7	2501	23	DNA encoding novel
21	285	53.6	2930	23	Drosophila melanog
22	285	53.6	3131	23	Drosophila melanog
23	284.5	53.5	2008	22	Human cDNA sequenc
24	284.5	53.5	2410	24	Nucleotide sequenc
25	284.5	53.5	3243	23	Human cDNA sequenc
26	284.5	53.5	3243	23	Genomic sequence #
27	284.5	53.5	3436	20	Murine IL-10 recep
28	284.5	53.5	4594	22	Human immune/haema
29	284	53.4	1110	22	Human immune/haema
30	284	53.4	3520	15	Interleukin-10 rec
31	283.5	53.3	3162	22	Human musculoskele
32	283	53.2	2533	19	Mouse gamma II ada
33	283	53.2	3520	19	Mouse IL-10 recept
34	282.5	53.1	1409	23	Human prostate exp
35	282.5	53.1	2342	12	Rat dopamine DI re
36	282.5	53.1	3025	12	D1 dopamine recept
37	282.5	53.1	3025	18	D1 dopamine recept
38	282	53.0	1001	21	Arachidonic acid m
39	282	53.0	3127	23	Drosophila melanog
40	281.5	52.9	1635	23	DNA encoding novel
41	281.5	52.9	4344	24	Gene #1754 used to
42	281.5	52.9	4388	18	Nucleotide sequenc
43	281	52.8	2693	14	NANBH virus strain
44	281	52.8	4168	21	Sequence encoding
45	280.5	52.7	1807	22	Human immune/haema

## ALIGNMENTS

### RESULT 1

AAD23462  
ID AAD23462 standard; cDNA; 337 BP.

XX AAD23462;

AC AAD23462;

DT 26-FEB-2002 (first entry)

DE Human lung tumour-specific 20E10 5' cDNA.

KW Human: lung tumour protein; immunostimulant; cytostatic; gene therapy;  
antisense-therapy; vaccine; immune response; lung cancer; 20E10; ss.

OS Homo sapiens.

PN WO200172295-A2.

XX 04-OCT-2001.

PD

XX 28-MAR-2001; 2001WO-US09991.  
 XX 29-MAR-2000; 2000US-0538037.  
 PR 05-JUN-2000; 2000US-0588937.  
 PR 18-AUG-2000; 2000US-0640878.  
 PR 22-SEP-2000; 2000US-234517P.  
 PR 01-NOV-2000; 2000US-0704512.  
 PR 14-DEC-2000; 2000US-0738973.  
 XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Mohamath R, Secrhist H, Benson DR, Indirias CY;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX WPI; 2001-639201/73.  
 XX New human lung-specific polynucleotides and polypeptides for the  
 PT diagnosis and treatment of disease e.g. lung cancer -  
 XX Claim 1; Page 334; 378pp; English.

XX The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer. The  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is a cDNA encoding human lung tumour-specific protein.

XX SQ Sequence 337 BP; 103 A; 60 C; 93 G; 81 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.4e-05 Length: 337  
 Score: 532.00 Matches: 97  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0

US-09-854-133-586 (1-97) x AAD23462 (1-337)

Qy 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 Db 5 GAGGTGAAGTGAGCAGAGATCATCCAGCCCTGGGTGACAGTGAGACTCTGTCTCAACA 64  
 Qy 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
 Db 65 GAATTAAGGAAAAAAGAAAGAAAGAAAGAGAGAGAGAAATTCAGGCCCAATTGTGGC 124  
 Qy 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHistrpIleGln 60  
 Db 125 ATAGATTTTATCATATCTGTGATTTTGGATTTTGGATTTTTCATCATCTGATTCAG 184  
 Qy 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 Db 185 GAAAGCCTGTGTGTCCACCATCTCCAAAGGAGGTACTCTCAGGGAATGTTAAGGGA 244  
 Qy 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArgArgLysCysSer 97  
 Db 245 GGCTGCCTTCCTGGGCAACAGGAGGACCTGGGAGGAGAAAGTGCGAGC 295

RESULT 2

ID AAD23460 standard; cDNA; 2239 BP.

XX AAD23460;

AC AAD23460;

XX 26-FEB-2002 (first entry)

XX Human lung tumour-specific 19A4 cDNA.

DE

XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 KW antisense-therapy; vaccine; immune response; lung cancer; 19A4; ss.  
 XX Homo sapiens.  
 XX WO200172295-A2.  
 XX 04-OCT-2001.  
 XX 28-MAR-2001; 2001WO-US09991.  
 XX 29-MAR-2000; 2000US-0538037.  
 PR 05-JUN-2000; 2000US-0588937.  
 PR 18-AUG-2000; 2000US-0640878.  
 PR 22-SEP-2000; 2000US-234517P.  
 PR 01-NOV-2000; 2000US-0704512.  
 PR 14-DEC-2000; 2000US-0738973.  
 XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Mohamath R, Secrhist H, Benson DR, Indirias CY;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX WPI; 2001-639201/73.  
 XX New human lung-specific polynucleotides and polypeptides for the  
 PT diagnosis and treatment of disease e.g. lung cancer -  
 XX Claim 1; Page 332; 378pp; English.

XX The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer. The  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is a cDNA encoding human lung tumour-specific protein.

SQ Sequence 2239 BP; 619 A; 444 C; 493 G; 683 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.0127 Length: 2239  
 Score: 523.00 Matches: 97  
 Percent Similarity: 84.35% Conservative: 0  
 Best Local Similarity: 84.35% Mismatches: 0  
 Query Match: 98.31% Indels: 18  
 DB: 23 Gaps: 1

US-09-854-133-586 (1-97) x AAD23460 (1-2239)

Qy 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 Db 2 GAGGTGAAGTGAGCAGAGATCATCCAGCCCTGGGTGACAGTGAGACTCTGTCTCAACA 61  
 Qy 21 GluLeuArgLysLysGluArgLysLysLysArgGluArgLysPheGlnAlaAsnCysGly 40  
 Db 62 GAATTAAGGAAAAAAGAAAGAAAGAAAGAGAGAGAGAAATTCAGGCCCAATTGTGGC 121  
 Qy 41 IleAspPheIleIlePheTrpIlePheTrpIleLeuLeuPheSerHisHistrpIleGln 60  
 Db 122 ATAGATTTTATCATATCTGTGATTTTGGATTTTGGATTTTTCATCATCTGATTCAG 181  
 Qy 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 Db 182 GAAAGCCTGTGTGTCCACCATCTCCAAAGGAGGTACTCTCAGGGAATGTTAAGGGA 241  
 Qy 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg----- 93  
 Db 242 GGCTGCCTTCCTGGGCAACAGGAGGACCCACCTGGGAGGAGCGCTTTTCAGGAAGAGACG 301

Pred. No.: 103 Length: 800  
Score: 339.50 Matches: 62  
Percent Similarity: 57.25% Conservative: 13  
Best Local Similarity: 47.33% Mismatches: 7  
Query Match: 63.82% Indels: 49  
DB: 20 Gaps: 28

US-09-854-133-586 (1-97) x AA216609 (1-800)

QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
DB 107 GAG---GAGGTGGAGAAATTGAGACGACGATCATACACAGGTGTTCTGAGTAGTAATTA 163  
QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
DB 164 ---GATCGCTGTGAAGGAAAGACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 217  
QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
DB 218 CGCTGAGAGACAGCTCTGAAAGCAGAGGAAAGACATCATAGTAACACCAAGACACACC 277  
QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49  
DB 278 AAA-----GTTGAAAGTTT---GTTTCTTTCCCTCTCTTTTATTT 316  
QY 50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGlnSerLeuLeuCysPro 66  
DB 317 TTCCCGCGTGTCTCTACTA-----TGG---TCAGAAAGCCTGTTGTGTCCA 361  
QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
DB 362 CCATCTCCAAAGGAGGTACCTGCGAGGAAATGTTAAACGGAGGCTGCCTTCCTCGGCA 421  
QY 87 ThrArgSerHisLeuGlyArgLysCysSer 97  
DB 422 ACAAGAGGCCACCTGGCGAGGAGAAAGTGCAGC 454

RESULT 4  
ABL37718  
ID ABL37718 standard; cDNA; 575 BP.

XX ABL37718;  
XX 08-APR-2002 (first entry)  
DT Human colon tumour antigen polynucleotide SEQ ID NO:1307.  
DE Human; colon cancer; colon tumour antigen; cytostatic; vaccine;  
KW colon tumour metastatic antigen; diagnosis; gene; ss.  
XX Homo sapiens.  
XX WO200196388-A2.  
XX 20-DEC-2001.  
XX 08-JUN-2001; 2001WO-US18557.  
XX 09-JUN-2000; 2000US-210899P.  
XX 20-FEB-2001; 2001US-270216P.  
XX (CORI-) CORIXA CORP.  
XX Jiang Y, Harlocker SL, Secrist H;  
XX WPI; 2002-114514/15.  
XX Novel isolated colon tumour polynucleotide differentially expressed in  
XX colon tumor or colon metastatic tumor and polypeptides encoded by them,  
XX useful for inhibiting development of cancer in patient -  
XX Claim 1; SEQ ID 1307; 105pp; English.

QY 94 -----ArgLysCysSer 97  
DB 302 CCTTTTCAGAGAGAGAGCGCTTTTTCAGGAGAGAGAAAGTGCAGC 346

RESULT 3  
AA216609  
ID AA216609 standard; cDNA; 800 BP.  
XX AA216609;  
DT 12-OCT-1999 (first entry)  
XX Human gene expression product cDNA sequence SEQ ID NO:4079.

XX Human; gene; gene expression product; diagnosis; therapy; probe;  
KW detection; mapping; tissue typing; profiling; forensic; cancer;  
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
XX Homo sapiens.  
XX WO9938972-A2.  
XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.  
XX 03-APR-1998; 98US-0080666.  
XX 28-JAN-1998; 98US-0072910.  
XX 24-FEB-1998; 98US-0075954.  
XX 31-MAR-1998; 98US-0080114.  
XX 03-APR-1998; 98US-0080515.  
XX (CHIR) CHIRON CORP.  
XX (HYSE-) HYSEQ INC.  
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
PI Lanson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
XX WPI; 1999-494092/41.  
XX Novel human genes and their expression products which are  
PT differentially expressed in different cell types

PS Claim 1; Page 1934; 2479pp; English.  
XX The present invention describes a library of human polynucleotides  
CC comprising the sequences given in AA212532 to AA21779. Also described is  
CC a method of detecting differentially expressed genes correlated with the  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample from a cell  
CC suspected of being cancerous, where the gene product is encoded by one  
CC of the 5248 polynucleotide sequences given in AA212532 to AA21779. The  
CC polynucleotides can be used as a source of primers and probes, which can  
CC be used for a variety of purpose, e.g. detection of expression levels,  
CC mapping, tissue typing or profiling, forensics, genetic analysis and  
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
CC can be used for raising antibodies for experimental, diagnostic and  
CC therapeutic purposes. The polynucleotides may also be used to construct  
CC arrays for diagnostics (which may be used to determine function of an  
CC encoded protein); and to detect differences in expression levels between  
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
CC identify a genetic predisposition or susceptibility to a disease such as  
CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
XX peptide analogues and antagonists.  
XX Sequence 800 BP; 199 A; 166 C; 202 G; 211 T; 22 other;  
SQ Alignment Scores:

CC ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)  
 CC which were isolated from human colon tumour and colon metastatic tumour  
 CC cDNA libraries. (I) have cytosolic activity and can be used in vaccine  
 CC production. (I) can be used for stimulating and/or expanding T cells  
 CC specific for a tumour protein on contact with the T cells. They are also  
 CC used for inhibiting the development of cancer in a patient. (I) can be  
 CC used as probes or primers for nucleic acid hybridisation, for preparing  
 CC mutant species primers, or primers for use in genetic constructions. (I)  
 CC can be used in the diagnosis of a colon tumour.

XX Sequence 575 BP; 151 A; 122 C; 153 G; 141 T; 8 other;

#### Alignment Scores:

Pred. No.: 36.2 Length: 575  
 Score: 339.00 Matches: 62  
 Percent Similarity: 57.25% Conservative: 13  
 Best Local Similarity: 47.33% Mismatches: 6  
 Query Match: 63.72% Indels: 50  
 DB: 24 Gaps: 28

US-09-854-133-586 (1-97) x ABL37718 (1-575)

QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
 Db 22 GAGGTGGAG-----AATTGAGACGACGATCATACACAGGTGTTCTGAGTAGTAATTA 75  
 QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
 Db 76 ---GATCGCTGTGAAGAAAGACACACCTTTGAGTTTACCTGTGAACA---CTATAG 129  
 QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
 Db 130 CGCTGAGAGACAGCTGTGAAGACAGAGAGACATCGATCAGTAACACCAAGACAC 189  
 QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49  
 Db 190 AAA-----GTTGAAGTTT---GTTTCTTCTTCCCTCTGTTTATT 228  
 QY 50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCyspro 66  
 Db 229 TTTCCCGGTGTGTCCTACTA-----TGG---TCAGAAAGCTGTGTGTCCA 273  
 QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
 Db 274 CCATCTCCAAAGAGGTTACCTGTCAGGAAATGTTAAACGGAGGCTGCCTTCCCTGGCA 333  
 QY 87 ThrArgSerHisLeuGlyArgArgLysCysSer 97  
 Db 334 ACAAGGACCCCTGGCAGGAGAAAGTGCAGC 366

#### RESULT 5

ABK27651/C  
 ID ABK27651 standard; cDNA; 596 BP.  
 AC ABK27651;  
 XX  
 XX  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Human colon cancer expressed sequence tag, Seq ID no 88.  
 XX  
 KW Human: colon cancer; T cell expansion; tumour; EST; gene; ss;  
 XX expressed sequence tag.  
 XX Homo sapiens.  
 OS  
 XX  
 XX  
 PN WO200196390-A2.  
 XX  
 XX 20-DEC-2001.  
 PD  
 XX 08-JUN-2001; 2001WO-US18577.  
 PF  
 XX 09-JUN-2000; 2000US-210821P.  
 PR 18-DEC-2000; 2000US-256571P.  
 PR

10-MAY-2001; 2001US-290240P.

(CORI-) CORIXA CORP.

Jiang Y, Hepler WT, Clapper JD, Wang A, Secrist H;  
 WPI; 2002-139708/18.

XX

Novel isolated polynucleotide encoding a polypeptide comprising a  
 portion of colon tumour protein, useful for detection, diagnosis and  
 therapy of human colon cancer

Claim 1; Page 174; 220pp; English.

The invention relates to an isolated polynucleotide (I) encoding a  
 polypeptide (II) comprising at least a portion of a colon tumour  
 protein. (I), (II) and antibody (III) to (II) are useful for determining  
 the presence of a cancer in a patient. (I), (II) or antigen presenting  
 cells expressing (I) is useful for stimulating and/or expanding T cells  
 specific for a tumour protein, by contacting T cells with (I), (II) or  
 antigen-presenting cells that express (I), under conditions and for a  
 time sufficient to permit the stimulation and/or expansion of T cells.  
 (I), (II), or antigen presenting cells that express (II) are useful for  
 treating colon cancer in a patient by incubating CD4+ and/or CD8+ T cells  
 isolated from a patient with (I), (II) or antigen presenting cells that  
 express (II), such that T cells proliferate, and administering to the  
 patient an effective amount of the proliferated T cells, thus inhibiting  
 the development of a cancer in the patient. (I) or (II) is useful in  
 vaccines and pharmaceutical compositions for prevention and treatment  
 of colon malignancies and for the diagnosis and monitoring of such  
 cancers. (I), (II) or (III) is useful for detection, diagnosis and/or  
 therapy of human colon cancer. (I) is useful as a probe or primer for  
 nucleic acid hybridisation, and in the design and preparation of  
 ribozyme molecules for inhibiting expression of (II) in tumour cells.  
 CC ABK27564-ABK27807 represent novel human colon cancer coding  
 CC sequences and primers of the invention.

XX Sequence 596 BP; 160 A; 151 C; 126 G; 158 T; 1 other;

#### Alignment Scores:

Pred. No.: 88.3 Length: 596  
 Score: 327.50 Matches: 59  
 Percent Similarity: 57.48% Conservative: 14  
 Best Local Similarity: 46.46% Mismatches: 5  
 Query Match: 61.56% Indels: 49  
 DB: 24 Gaps: 27

US-09-854-133-586 (1-97) x ABK27651 (1-596)

QY 6 Arg-----AspHis---Ala-----Ser-----LeuGlyAsp---Ser 14  
 Db 595 AAGAAATTGAGAGCAGCATGCATACACAGGTCTTCTGAGTAGTAATTA---GATCGCTGT 539  
 QY 15 Glu-----Thr---LeuSer---Gln-----ThrGluLeu---Arg-----Lys 24  
 Db 538 GAAGNAAAAAGACACACCTTTGAGTTTCACTGTGAACA---CTATAGCGCTGAGAGAGA 482  
 QY 25 Lys---Glu---Arg---Lys---LysLys-----ArgGluArgLysPheGlnAla 37  
 Db 481 CAGTCTGAAAGCAGAGAGAGACATCGATCAGTAACACCAAGACACCAAA----- 431  
 QY 38 AsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---PheTrp-----Ile 51  
 Db 430 -----GTTGAAAGTTT---GTTTCTTCTTCCCTCTGTTTATTTTCCCGCGGTG 383  
 QY 52 ----LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysProSerProLys 70  
 Db 382 TCCCTACTA-----TGG---TCAGAAAGCTGTGTGTCCCATCTCCAAAG 338  
 QY 71 GluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAlaThrArgSerHis 90  
 Db 337 GAGTTACCTGCAGGGAATGTTAAAGGGAGGCTGCTTCCCTGGGCAACAGAGGCAC 278

QY	91	LeuGLYArGArGLysCysser	97
DB	277	CTGGCGAGGAGAAATGTCAGC	257
	RESULT 6		
	ABAL6566/c		
ID	ABAL6566	standard; DNA; 2752 BP.	
XX	AC		
XX	ABAL6566;		
XX	23-JAN-2002	(first entry)	
XX	Human	nervous system related pol	
XX	Human;	nootropic; neuroprotectiv	
KW	Human;	nootropic; neuroprotectiv	
KW	immunosuppressive;	antiinflammat	
KW	antiparkinsonian;	antiskinfat	
KW	antirheumatic;	hepatotropic; anti	
KW	antiallergic;	antidiabetic; anti	
KW	antiparasitic;	cardiant; immune	
KW	neurological disease;	infection;	
XX			
XX	Homo sapiens.		
XX	WO200159063-A2.		
PN	16-AUG-2001.		
XX			
PD	17-JAN-2001;	2001WO-US01334.	
PF			
XX	31-JAN-2000;	2000US-0179065.	
XX	04-FEB-2000;	2000US-0180628.	
PR	24-FEB-2000;	2000US-0184664.	
PR	02-MAR-2000;	2000US-0186350.	
PR	16-MAR-2000;	2000US-0189874.	
PR	17-MAR-2000;	2000US-0190076.	
PR	18-APR-2000;	2000US-0198123.	
PR	19-MAY-2000;	2000US-0205515.	
PR	07-JUN-2000;	2000US-0209467.	
PR	28-JUN-2000;	2000US-0214886.	
PR	30-JUN-2000;	2000US-0215135.	
PR	07-JUL-2000;	2000US-0216647.	
PR	07-JUL-2000;	2000US-0216880.	
PR	11-JUL-2000;	2000US-0217487.	
PR	11-JUL-2000;	2000US-0217496.	
PR	14-JUL-2000;	2000US-0218290.	
PR	26-JUL-2000;	2000US-0220963.	
PR	26-JUL-2000;	2000US-0220964.	
PR	14-AUG-2000;	2000US-0224518.	
PR	14-AUG-2000;	2000US-0224519.	
PR	14-AUG-2000;	2000US-0225213.	
PR	14-AUG-2000;	2000US-0225214.	
PR	14-AUG-2000;	2000US-0225266.	
PR	14-AUG-2000;	2000US-0225267.	
PR	14-AUG-2000;	2000US-0225268.	
PR	14-AUG-2000;	2000US-0225270.	
PR	14-AUG-2000;	2000US-0225447.	
PR	14-AUG-2000;	2000US-0225757.	
PR	14-AUG-2000;	2000US-0225758.	
PR	14-AUG-2000;	2000US-0225759.	
PR	18-AUG-2000;	2000US-0226279.	
PR	22-AUG-2000;	2000US-0226681.	
PR	22-AUG-2000;	2000US-0226868.	
PR	22-AUG-2000;	2000US-0227182.	
PR	23-AUG-2000;	2000US-0227009.	
PR	30-AUG-2000;	2000US-0228924.	
PR	01-SEP-2000;	2000US-0229287.	
PR	01-SEP-2000;	2000US-0229343.	
PR	01-SEP-2000;	2000US-0229344.	
PR	01-SEP-2000;	2000US-0229345.	
PR	05-SEP-2000;	2000US-0229509.	
PR	05-SEP-2000;	2000US-0229513.	
PR	06-SEP-2000;	2000US-0230437.	



us-09-854-133-586.rng

Tue May 13 12:12:49 2003

PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216547.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225271.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234597.  
 PR 25-SEP-2000; 2000US-0234598.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246539.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2000US-0254097.  
 PR 05-JAN-2001; 2000US-02559678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-488786/53.  
 XX New isolated ovarian and/or breast cancer related nucleic acids and  
 XX polypeptides, useful for diagnosing, treating and/or preventing human  
 XX diseases and disorders, particularly ovarian and/or breast cancer -  
 PS Disclosure; SEQ ID NO 711; 577pp + Sequence Listing; English.  
 XX The invention relates to novel genes (ABA07454-ABA08224) and proteins  
 CC (ABB10743-ABB10980) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are

\_\_\_\_\_



PR	14-SEP-2000;	2000US-0233063.
PR	14-SEP-2000;	2000US-0233064.
PR	14-SEP-2000;	2000US-0233065.
PR	14-SEP-2000;	2000US-0233066.
PR	21-SEP-2000;	2000US-0234223.
PR	21-SEP-2000;	2000US-0234224.
PR	21-SEP-2000;	2000US-0234227.
PR	25-SEP-2000;	2000US-0234297.
PR	25-SEP-2000;	2000US-0234998.
PR	25-SEP-2000;	2000US-0235484.
PR	26-SEP-2000;	2000US-0235834.
PR	27-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249266.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249269.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	06-DEC-2000;	2000US-0251856.

<b>Alignment Scores:</b>					
Pred. No.:	1.88e+04	Length:	1687		
Score:	298.50	Matches:	58		
Percent Similarity:	39.78%	Conservative:	16		
Best Local Similarity:	31.18%	Mismatches:	10		
Query Match:	56.11%	Indels:	102		
DB:	22	Gaps:	38		
US-09-854-133-586 (1-97) x AAL03733 (1-1687)					
QY	1	GluValGlu-----valSerArgAspHisala-----SerLeu-----Gly-----AspSer	14		
Db	633	GAGGTAGAGTGGCAGTCAGTCAGTCGAGATCACGCCACTGCCTTCACCTCCAGCGTGCGTGCAGACAGC	692		
QY	15	GluThrLeuSerGlnThrGluLeuArgLysLysGlu-----ArgLys-----Lys	29		
Db	693	AAGACTTTGTCTCA-----AAAAAAAAAAGACCTACTCTATAGGAAGAACAATTACAG	743		
QY	30	LysArg---GluArgLysPheGlnIalaasn-Cys-----Gly-----	40		
			:::		
Db	744	AAAAGAGTAAC-----AACTATATG-----ATAATGTTAGTTTCAGTAGTAATCAGGACAGTG	797		
QY	41	----ile-----AspPheIleile-----PheTrp-----Ileph	49		
			:::		:::
Db	798	CAAATCAAATACAGTCAGAT---CAAGTGCTCCTTATTGGTAAAAAATTAAAACTTT	854		
QY	49	eTrpIleLeuLeu-----PheSer-----His---Histrp-----Ilegl	60		
		:::			
Db	855	TITTTTTCTTTTTTGGAGATGCCTTCTCACTTTGTCCACCAGGTGGAGTGCAGTGGTGCA	914		
QY	60	nLusEr---LeuLeu-----LysProSer-----CysProProSer---	68		
Db	915	A---TCTGGCTCATTTGGAACCTCTCCCTCCCAGGCTCAATCAGTCTGCCACCTCAGCC	971		
QY	69	-ProLys-----GluVal-----Thr-----Cys-----	74		
			:::		
Db	972	TCCCAAGTAGCTGTGACTACAGTCTCTCCCAACAGCGCTGGCTAAATTTCTTGGCATTTT	1031		
QY	75	---ArgGluMetLeuThrGly-----GlyCysLeupro---TriAla---Thr--	87		
		:			:::
Db	1032	TGATAGAGAT-----GGAGTTTCGCCAGGTTCG---CCAAGCTGGTCTCAAACTCC	1079		
QY	88	-----Arg---Ser---HisLeuGly-----Argarg-----	94		
		:::			
Db	1080	TGAGCTCAAGCATCCGCCACCTTGGCCCTCCCAAACTACTGGGCTTTACAGGGGTGAGCC	1139		
QY	95	-----Lys	95		

Db 1140 ACCGGACCCACGACAAA 1155  
RESULT 9  
ABAO7917  
ID ABAO7917 standard; DNA; 1688 BP.  
XX AC ABAO7917;  
XX DT 11-JAN-2002 (first entry)  
XX DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 712.  
XX KW Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein; ds.  
XX OS Homo sapiens.  
XX PN WO200155325-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01345.  
XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216847.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225271.  
PR 14-AUG-2000; 2000US-0225275.  
PR 14-AUG-2000; 2000US-0225278.  
PR 18-AUG-2000; 2000US-0226379.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 25-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 26-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 02-OCT-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-024617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.

Tue May 13 12:12:49 2003

```

PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-488786/53.
XX
XX New isolated ovarian and/or breast cancer related nucleic acids and
XX polypeptides, useful for diagnosing, treating and/or preventing human
XX diseases and disorders, particularly ovarian and/or breast cancer -
XX
XX Disclosure; SEQ ID NO 712; 577pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA07454-ABA08224) and proteins
XX (ABB10743-ABB10980) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus and ulcerative
XX disease, multiple sclerosis, rheumatoid arthritis and Crohn's
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1688 BP; 542 A; 332 C; 358 G; 456 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.89e+04 Length: 1688
XX Score: 298.50 Matches: 58
XX Percent Similarity: 39.78% Conservative: 16
XX Best Local Similarity: 31.18% Mismatches: 10
XX Query Match: 56.11% Indels: 102
XX DB: 22 Gaps: 38
XX
XX US-09-854-133-586 (1-97) x ABA07917 (1-1688)
XX
XX QY 1 GluValGlu-----ValSerArgAspHisAla---SerLeu-----Gly---AspSer 14
XX Db 634 GAGTGGAGGTTGCGATGATCGAGATCAGCCACCTGCCTTCAGCTGGTGACAGAGC 693
XX
XX QY 15 GluThrLeuSerGlnThrGluLeuArgLysLysGlu-----ArgLys-----Lys 29
XX Db 694 AAGACTTTGTCTCAA-----AAAAAAGACCTACTCTATAGGAAGCAATTCAG 744
XX
XX QY 30 LysArg---GluArgLysPheGlnAlaAsn-Cys-----Gly----- 40
XX Db 745 AAAAGAGTAAAC-----AACTATATG---AATATGTTAGTTTCAGTAGTAATCAGGACGTG 798

```

```

QY 41 ----Ile-----AspPheIleIle-----PheTrp-----IlePh 49
Db 799 CAAATCAAAATACAGTGAGAT---CAAGTGTCTCTTATTTGTTAAATAAATTTT 855
QY 49 etrPileLeuLeu-----PheSer-----His---HisTrp-----IleGl 60
Db 856 TTTTTCCTTTTGGAGATGCTCTCTCACTTTGTACCCAGGTTGGAGTGCAGTGTGCA 915
QY 60 nGluSer---LeuLeu-----CysProProSer--- 68
Db 916 A---TCTGGCTCATTTGCAACCTCTGCCTCCAGGCTCAATCACTGTGCCCACTCAGCC 972
QY 69 -ProLys-----GluVal-----Thr-----Cys----- 74
Db 973 TCCCAAGTAGCTGTGATACAGTCTCCACCACGCTGCTGCTTAATTTCTTTGCAATTT 1032
QY 75 ---ArgGluMetLeuThrGly-----GlyCysLeuPro---TrpAla---Thr--- 87
Db 1033 TGATAGAGAT-----GGAGTTTGGCCAGGTTGC---CCAAGTGGTCTCAAACTCC 1080
QY 88 -----Arg---Ser---HisLeuGly-----ArgArg----- 94
Db 1081 TGAGCTCAAGCAATCCGCCACCTTGCTCCCAAGTACTGGCTTACAGGCGTGAGCC 1140
QY 95 -----Lys 95
Db 1141 ACCGGACCCAGCAAAA 1156
XX
XX RESULT 10
XX AAL03734
XX ID AAL03734 standard; DNA; 1688 BP.
XX AC AAL03734;
XX DT 21-NOV-2001 (first entry)
XX DE Human reproductive system related antigen DNA SEQ ID NO: 6422.
XX KW Human; reproductive system related antigen; reproductive system disorder;
XX OS Homo sapiens.
XX PN WO200155320-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01339.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.

```

1 GluValGlu-----ValSerArgAspHisAla-----SerLeu-----Gly---AspSer 14  
|||||  
634 GAGTGGAGGTTGCAGTGTAGCTCAGATCATCGCCACTGTCAGCTTCAGCCTGGGTGACAGAGC 69

15 GluThrLeuSerGlnThrGluLeuArgLysLysGlu-----ArgLys-----Lys 29  
:::|||||  
694 AAGACTTTGTCTCAA-----AAAAAAGACCTACTCTATAGGAAGCAATTCAG 744  
30 LysArg-----GluArgLysPheGlnAlaAsn-Cys-----Gly----- 40  
|||||  
745 AAAAGAGTAAC-----AACTATATG---AATATGTTAGTTTCAGTAGTAATCAGGACAGTG 798  
41 -----Ile-----AspPheIleIle-----PheTrp-----IlePhe 49  
|||||  
799 CAAATTCAAATAACAGTGAGAT---CAAGTGTCTCCTTATTTGGTAAAAATTAAACTTTT 855  
49 enrPileLeuLeu-----PheSer-----His---HisTrp-----IleG1 60  
856 TTTTCTTTTGTGGAGATGCCCTTCTCACCCTTGTGCACCCAGGTTGGAGTGCAGTGGTGCA 915  
60 nGluSer---LeuLeu-----CysProProSer--- 68  
916 A--TCTGGCTCATTTGCAACTCTGCTCCAGGCTCAATCACTGTGCCACCTCAGCC 972  
69 -ProLys-----GluVal-----Thr-----Cys----- 74  
973 TCCCAAGTAGCTGTGACFACAGGTCTCTCCACCACGCTGGCTAAATTTTCTTGCATTTT 1032  
75 ----ArgGluMetLeuThrGly-----GlyCysLeuPro---TrpAla---Thr-- 87  
1033 TGATAGAGAT-----GGAGTTGCGCAGGTTGC---CCAAGCTGGTCTCAAACTCC 1080  
88 -----Arg---Ser---HisLeuGly-----ArgArg----- 94  
1081 TGAGCTCAAGCAATCGCCCACTTGGCTCCCAAAGTACTGGCTTACAGGCGTGAGCC 1140  
95 -----Lys 95  
1141 ACCGGACCCAGCAAAA 1156  
RESULT 11  
C BK35489  
X BK35489 standard; DNA; 3995 BP.  
X BK35489;  
X 08-MAY-2002 (first entry)  
X Human endometrial cancer related gene, DLG4.  
X Human; ds; gene; endometrial cancer; differential expression;  
X DNA microarray; protein microarray.  
X Homo sapiens.  
X WO200209573-A2.  
X 07-FEB-2002.  
X 31-JUL-2001; 2001WO-US24104.  
X 31-JUL-2000; 2000US-221735P.  
X (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
X Mutter GL;  
X WPI; 2002-179967/23.  
X P-PSDB; AAU84269.  
X Diagnosing endometrial cancer comprises determining expression of  
X nucleic acid molecules or expression products that are differentially  
X expressed in normal and malignant endometrium -  
X Claim 1; Page 56-58; 233pp; English.  
X The invention relates to diagnosing endometrial cancer in a subject

the invention relates to diagnosing endometrial cancer in a subject

QY 77 MetLeu-----Thr-----GlyGly-----Cys--- 82  
 Db 3350 ---CTGGGCTCTGAGCCAGCAACCTGGGGTCTGGGGGAGCTGGGCTCTGTTC 3406  
 QY 83 -----Leu-----ProTriaPala----- 86  
 Db 3407 GAGCCCTGTCTCTTAGGATCCCG---GCCCCACCTGGCCCAATGCACACAGACCC 3463  
 QY 87 -----His-----Thr-----ArgSer----- 89  
 Db 3464 ACCGGGGGCACCTGCCCTCCCGCATCTCTCCACACATTCACAGAGTCAAGGGCCCC 3523  
 QY 90 -----Glu-----Gly-----Arg----- 94  
 Db 3524 CTGAGGAGCAGCCCGCTGAGGATCCAGGGCCACAGGCTCGCTCTCTCTTAAGGCAG 3583  
 QY 95 -----Lys-----Cys-----Ser 97  
 Db 3584 GGTCTGGGTACCCCTGCTCATGCTATTCCTCCCATGTTACCTTGATTCTCA 3637

## RESULT 12

ABK35548

ID ABK35548 standard; DNA; 3995 BP.

XX AC

XX ABK35548;

XX DT

XX 08-MAY-2002 (first entry)

XX DE

XX Gene DLG4 differentially expressed in breast cancer tissue.

XX KW

XX Human; diagnosis of breast cancer; endometrial cancer; breast tumour;

XX KW

XX MAI; mitotic activity index; cytostatic; gene; ds.

XX OS

XX Homo sapiens.

XX PN

XX WO200210436-A2.

XX PD

XX 07-FEB-2002.

XX 27-JUL-2001; 2001WO-US23542.

XX 28-JUL-2000; 2000US-222093P.

XX PA (BGHM) BRIGHAM &amp; WOMENS HOSPITAL INC.

XX PA (BAAK/) BAAK J.

XX PI

XX Baak J, Mutter GL;

XX DR WPI; 2002-180084/23.

XX DR P-PSDB; AAD04328.

XX PT

XX Diagnosing breast cancer comprises determining expression of nucleic

XX PT acid molecules or expression products that are differentially expressed

XX PT in normal and malignant tissue

XX FS

XX Claim 1; Page 58-61; 219pp; English.

XX CC

XX The present invention relates to a method for diagnosing breast cancer

XX CC in a subject suspected of having endometrial cancer. The method

XX CC comprises determining the expression of a set of human genes or

XX CC expression products in an endometrial sample suspected of being

XX CC cancerous. The human genes of the invention are differentially

XX CC expressed in breast tumours characterised as high or low MAI (mitotic

XX CC activity index). These sets of genes can be used to discriminate between

XX CC high and low MAI breast tumours. The invention also provides DNA and

XX CC protein microarrays for analysing the expression of the human genes and

XX CC their protein products. The methods and arrays are useful for the

XX CC diagnosis and prognosis of endometrial cancer, selecting and monitoring

XX CC treatment regimes, and identification of compounds useful for the

XX CC treatment of endometrial cancer. ABK35531-ABK35581 represent the human

XX CC genes of the invention that are differentially expressed in breast

XX CC cancer tissue.

XX CC

SQ Sequence 3995 BP; 901 A; 1113 C; 1160 G; 821 T; 0 other;

## Alignment Scores:

Pred. No.: 4.5e+05 Length: 3995  
 Score: 292.50 Matches: 69  
 Percent Similarity: 24.53% Conservative: 9  
 Best Local Similarity: 21.70% Mismatches: 11  
 Query Match: 54.98% Indels: 229  
 DB: 24 Gaps: 68

US-09-854-133-586 (1-97) x ABK35548 (1-3995)

QY 1 GluVal-----Glu-----Val-----SerArg-----Asp--- 7  
 Db 2708 GAGATAGATGGCGGGATTAACCACTTTGTGTCTCGCGGAGAAAATGGAGAGACATT 2767  
 QY 8 -----His-----Ala-----Ser-----Leu-----Gly----- 12  
 Db 2768 CAGGGCCACAAAGTTTCAATGAGCGCGCCAGTACACAGCCACCTCTATGGACACAGGTC 2827  
 QY 13 AspSer-----Glu-----Thr-----Leu-----Ser----- 18  
 Db 2828 CAGTCCGTGCGAGAGGTGCGACAGCAGGGGAAGCACTGCATCTCGGCCAAT 2887  
 QY 19 -----Gln-----ThrGluLeu-----Arg-----Lys 24  
 Db 2888 GCCGTGCGCGGCTGCAGCGCGGCCACCTGCACCCCATCGCCATCTTCATCGCCGCCCGC 2947  
 QY 25 Lys-----GluArg-----Lys-----LysArg-----Glu-----ArgLys--- 34  
 Db 2948 TCCCTGGAGATGTCTAGAGATTAAACAGCGGATCAACAGAGGAGCAAGCCGCAAGCC 3007  
 QY 35 Phe-----Gln-----AlaAsnCys----- 39  
 Db 3008 TTCACAGAGCCACCAAGCTGGAGCAGGAGTTACAGAGTCTTCTCAGCCATCGTGAG 3067  
 QY 40 GlyIleAsp-----Phe-----Ile-----Ile----- 45  
 Db 3068 GGT-----GACAGCTTTGAGAGATCTACCACAAAGTGAAGCGTGCATCAGGAGACTCTCA 3124  
 QY 46 -----Phe-----TrpIle-----Phe-----TrpIle-----Leu 52  
 Db 3125 GGCCCTTACATCTGGGTTCAGCCCGAGAGAGACTCTGATTCCTGCCCTGGCTGGCTG 3184  
 QY 53 -----LeuPheSerHisHis-----TrpIleGlnGluSer-----LeuLeu 64  
 Db 3185 GACTCGCCCTGCCTC-----CATCACTGGGCGCTGG-----TCTGACTGAAT 3229  
 QY 65 Cys-----ProPro-----Ser-----Pro----- 69  
 Db 3230 TGCCCAAGCCCTGGCTGCCCGCGCGCTCCCTCCACCCCTTCTTATTTATTTCTTCT 3289  
 QY 70 Lys-----Glu-----Val-----Thr-----Cys-----Arg-----Glu 76  
 Db 3290 AACTGGATCCAGCCCTGTTGGAGGGGGGACACTCTCTCGATGTATCCCGCACCAGAA 3349  
 QY 77 MetLeu-----Thr-----GlyGly-----Cys--- 82  
 Db 3350 ---CTGGCTCCTGAACCCAGGAACCTGGGTCTGGGGGAGCTGGGCTCTGTTC 3406  
 QY 83 -----Leu-----ProTriaPala----- 86  
 Db 3407 GAGCCCTTGTCTCTTAGATCCCG---GCCCCACCTGCCCAATGCACACAGACCC 3463  
 QY 87 -----His-----Thr-----ArgSer----- 89  
 Db 3464 ACCGGGGGCACCTGCCCTCCCGCATCTCTCCACACATTCACAGAGTCAAGGGCCCC 3523  
 QY 90 -----His-----Leu-----Gly-----Arg----- 94  
 Db 3524 CTGAGGAGCAGCCCGTGCAGGATGCAGGGCCACAGGCTCGCTCTCTCTTAAGGCAG 3583  
 QY 95 -----Lys-----Lys-----Cys-----Ser 97

us-09-854-133-586.rng

Tue May 13 12:12:49 2003

Db 3584 GGTCGGGTACCCCTGCCTCATGTAATCCCATGTTACCTGTGATTTCTCA 3637

RESULT 13  
ABA07287/c  
ID ABA07287 standard; DNA: 2197 BP.

XX ABA07287;

XX 14-JAN-2002 (first entry)

XX Human pancreatic cancer related genomic DNA, SEQ ID NO: 606.

XX Human; cytostatic; antidiabetic; antiinflammatory; gastric; osteopathic;  
XX antihormone; antiulcer; thyroid-active; gene therapy; antisense therapy;  
XX pancreatic cancer antigen inhibitor; pancreatic cancer; pancreatitis;  
XX diabetes; endocrine disorder; acromegaly; hyperthyroidism;  
XX gastrointestinal disorder; Crohn's disease; duodenal ulcer; ds.

XX Homo sapiens.

XX WO200155206-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01353.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 27-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.

Db	1781	TCCTCTGGGAGCAGACACTGAACAGGGCAGACCAAGGTTCTACCGAAGAAGGTGGTT	1722
QY	49	-----PheTrpIleLeuLeuPheSerHisHis-----	TripleGln---Glu 61
Db	1721	CCATGTTCTGG-----CCATTTTCC-----CATGGAGTGTGCACCTGG---GAGGTGAT	1674
QY	62	Ser-----Leu-----LeuCys-----ProProSerProLysGlu-----	Val 72
Db	1673	TCATTGGACCCCTGGGGGTTGTCTGGGTGCTCAGCGAGCCACGAGGGCAGAGCTCAGTG	1614
QY	73	-----ThrCys-----ArgGluMet-----LeuThr-----Gly-----	Gly--- 81
Db	1613	AGCACTGTCTGGTGGAGAGCACTGAACACCAAGTCTCAGCCACGAGCCAGGATGGGGCC	1554
QY	82	-----Cys-----Leu-----Pro-----Trp-----	Trp--- 85
Db	1553	CAGGCAGAACCCACAGCCGATGCCCCCTCTCCTGTCGCCACCACTTCCTCTGTGTCT	1494
QY	86	-----AlaThr-----Arg-----Ser-----	Ser----- 89
Db	1493	CGGTACCCCTCTACGCACCTCAGGCTCGAGAAGCAAGCTGACCTCAGGCGCTAGACCCT	1434
QY	90	-----His-----Leu-----Gly-----Arg 93	Arg 93
Db	1433	GCAGGGAATGCTCTTCAAGATCATGACCTCCTTAATGATAAATTTGAACGGGCCCAA	1374
QY	94	ArgLys---CysSer 97	
Db	1373	---AAACAAGCATCC 1362	
RESULT 14			
AAK89932/c			
ID	AAK89932 standard; DNA; 2197 BP.		
XX	AAK89932;		
AC			
XX	05-NOV-2001 (first entry)		
DT	Human digestive system antigen genomic sequence SEQ ID NO: 3508.		
DE	Human; digestive system antigen; gene therapy; cancer; appendicitis;		
XX	ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;		
KW	digestive system disorder; Meckel's diverticulum; ds.		
KW			
OS	Homo sapiens.		
XX			
PN	WO200155314-A2.		
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01324.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	11-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	14-JUL-2000; 2000US-0217496.		
PR	26-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	14-AUG-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		



Alignment Scores:	6.88e+04	Length:	2197
Pred. No.:	292.00	Matches:	59
Score:	36.22%	Conservative:	18
Percent Similarity:	26.22%	Mismatches:	12
Best Local Similarity:	54.89%	Indels:	136
Query Match:	22	Gaps:	49

us-09-854-133-586.rng

US-09-854-133-586 (1-97) x AAK89932 (1-2197)

Qy	1	GluValGlu-----ValSerArgAspHis-----AlaSerLeu-----Gly---Asp 13
Db	2012	GAGCGGGAGTTGGCAGTGGCGGGAGATTCACACCATTCGA-----CTCCAGCGTGGCGCAACAG 1956
Qy	14	SerGluThrLeu-----SerGlnThrGluLeuArgLysLysGluArg---LysLys--- 29
Db	1955	AGTGAGACTCTGCCTTAAAAAATA-----AAA-----AAAAAAGAAAGCTGACAGAAGACA 1902
Qy	30	Lys-----Arg-----GluArg-----Lys---Phe-----GlnAla 37
Db	1901	GAACGTACAGGAAGTAGAAGTGCACCTGTGAAGAAAGGAGATAAATGCATATAAGACAGGCC 1842
Qy	38	---Asn-----Cys-----GlyIleAsp-----PheIle----- 44
Db	1841	CTAATGGCTGAAGAATTAGGACAGGCTGCTGTCCAGGAAGTGACAGATGGCTGTGCTG 1782
Qy	45	---IlePhe-----Tyr-----Lys-----TrpIle 48
Db	1781	TCCTCTCTGGGAGCAGACACTGACAGGGCAGACCAAGGTTCTACCGAAGAGGTGGGTT 1722
Qy	49	-----PheTrpIleLeuLeuPheSerHisHis-----TrpIleGlnb---Glu 61
Db	1721	CCATGTTCTGG-----CCATTTTCC-----CATGGAGTGTGCACCTGG---GAGGCTGAT 1674
Qy	62	Ser-----Leu---LeuCys-----ProProSerProLysGlu-----Val 72
Db	1673	TCTTGGACCTCGGGGTTTGTCTGGGTGCTGTACGCGAGCCACGAGGCGAAGCTCAGTG 1614
Qy	73	---ThrCys-----ArgGluMet-----LeuThr---Gly--- 81
Db	1613	AGCACTTGCTGTGTGAGAGCACTGAACCAAGTCTCAGCCAGGAGCCAGGATGGGCC 1554
Qy	82	-----Cys-----Leu---Pro-----Trp--- 85
Db	1553	GAGGCCAGAACCCACACAGCCGATGCCCTCTCTCTGTGCCACCACTTCCTGTGGTCT 1494
Qy	86	-----AlaThr-----Arg-----Ser----- 89
Db	1493	CGGTCAACCTCTACGCACTCAGGCTGCAGAGAGGAGGAGCTGACCTCAGGCTAGACCT 1434
Qy	90	-----His-----Leu---Gly---Arg 93
Db	1433	CGAGGGAATGCTCTTCCAGATCATGACCTCTCTAATCGATAAATTTGACGGGCCCAAA 1374
Qy	94	ArgLys---CysSer 97
Db	1373	---AAACAAGCATCC 1362
RESULT 15		
ABAJ16568/c		
ID	ABAJ16568	standard; DNA; 2751 BP.
XX	XX	ABAJ16568;
AC	XX	
XX	XX	
DT	DT	23-JAN-2002 (first entry)
DE	DE	Human nervous system related polynucleotide SEQ ID NO 8899.
XX	XX	Human; nontropic; neuroprotective; cytostatic; dermatological; virocid;
KW	KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerar;
KW	KW	antiParkinsonian; antiscikling; antianemic; antiarthritic; cancer;
KW	KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW	KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW	KW	antiparasitic; cardiatic; immune disorder; cardiovascular disorder;
XX	XX	neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
OS	OS	Homo sapiens.
PN	PN	WO200159063-A2.
XX	XX	
PD	PD	16-AUG-2001.



Search completed: May 11, 2003, 18:48:35  
Job time : 161 secs

Tue May 13 12:12:50 2003

us-09-854-133-586.rni

Sequence 26, Appl  
Sequence 3, Appl  
Sequence 24, Appl  
Sequence 1, Appl  
Sequence 24, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 58, Appl  
Sequence 58, Appl  
Sequence 58, Appl  
Sequence 995, Appl  
Sequence 3, Appl  
Sequence 1, Appl  
Sequence 28, Appl  
Sequence 28, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 86, Appl  
Sequence 167, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 2, Appl  
Sequence 302, Appl  
Sequence 1, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 7, Appl

US-09-018-635-26  
US-09-467-642-3  
US-09-041-886-24  
PCT-US94-05277-1  
US-09-724-864-24  
US-09-249-697A-2  
US-09-363-316B-2  
US-08-631-607-1  
US-09-098-358B-1  
US-09-232-201-58  
US-09-232-197-58  
US-09-232-200-58  
US-09-232-201-58  
US-09-221-017B-995  
US-08-106-761-3  
US-09-436-983-1  
US-08-371-930-28  
PCT-US94-01712-28  
US-09-013-895A-1  
US-09-565-318-1  
US-09-448-868-1  
US-09-058-489-86  
US-09-221-017B-167  
US-08-138-446B-9  
US-08-870-693-9  
US-09-228-986-2  
US-09-149-476-302  
US-08-824-996-1  
US-09-999-811-3  
US-09-042-105-3  
US-08-999-811-1  
US-09-042-105-1  
PCT-US96-09001-1  
US-08-795-430-7

ALIGNMENTS

RESULT 1  
US-08-861-745B-2/c  
Sequence 2, Application US/08861745B  
Patent No. 6165733

GENERAL INFORMATION:  
APPLICANT: Cen, Hul  
APPLICANT: Williams, Lewis  
TITLE OF INVENTION: Gamma II Adaptin  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Wilcoff  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/861,745B  
FILING DATE: 22-MAY-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A  
REGISTRATION NUMBER: 32141  
REFERENCE/DOCKET NUMBER: 02441.05336  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX:

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 18:46:17 ; Search time 42 Seconds  
(without alignments)  
708.277 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEYSRDBASLGDSLTSLT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62 0.0, Xgapext 0.5  
Xgapop 0.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues  
Total number of hits satisfying chosen parameters: 875442

Minimum DB seq length: 0  
Maximum DB seq length: 5000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09854133/runat\_05052003.174426.1235/app\_query.fasta\_1.263  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=5000  
-USER=US09854133 -CGN=1.1.32 -runat\_05052003.174426.1235 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=0 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	283	53.2	2533	4	US-08-861-745B-2
2	283	53.2	3520	1	US-08-424-788-1
3	283	53.2	3520	1	US-08-110-683-3
4	283	53.2	3520	2	US-08-477-166-3
5	283	53.2	3520	2	US-08-472-097-3
6	283	53.2	3520	2	US-09-439-672-3
7	283	53.2	3520	5	PCT-US93-11638-3
8	282.5	53.1	3025	1	US-08-444-734A-1
9	282	53.0	1001	4	US-09-641-638-319
10	281	52.8	2693	1	US-07-925-695-4
C 11	279.5	52.5	2907	2	US-09-018-638-17
C 12	279.5	52.5	2907	3	US-09-273-378-17



us-09-854-133-586.rnl

Tue May 13 12:12:50 2003

CLASSIFICATION: 436  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/011,066  
 FILING DATE: 29-JAN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ching, Edwin P.  
 REGISTRATION NUMBER: 34,090  
 REFERENCE/DOCKET NUMBER: DX0335K1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-496-1200  
 TELEFAX: 415-496-1200  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3520 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 80..1807  
 US-08-110-683-3

Alignment Scores:  
 Pred. No.: 4.87e+04  
 Score: 283.00  
 Percent Similarity: 25.66  
 Best Local Similarity: 20.72  
 Query Match: 53.20  
 DB: 1  
 Length:  
 Matches: 3520  
 Conservative: 63  
 Mismatches: 15  
 Indels: 216  
 Gaps: 62

US-09-854-133-586 (1-97) x US-08-110-683-3 (1-3520)

QY 1 Glu-----ValGluValSerArg-----Asp-----HisAla 9  
 DB 2001 GAGTGGGCTCTGTAGTACACGAGCAGCTGAGCAGGATTGACAGAGACCTCTCTCATGCC 2060  
 QY 10 Ser-----Leu-----Gly----- 12  
 DB 2061 TCAGGGCTGGCTCTACACTGGAAGACCTGTGTGGGTGTACCTCAGGCTTTCTGG 2120  
 QY 13 -----Asp-----SerGlu-----Thr-----LeuSer-----GlnThr-----Glu 21  
 DB 2121 ATGTGTAAGACTGTAGTCTGAGTCTGAGCTGAGCTGAGCTGGATCTCTGCGGAGGTGTGGAG 2180  
 QY 22 -----Leu-----Arg-----Lys-----LysGluArg-----Lys-----Lys----- 29  
 DB 2181 TGGTAGCTGCTACAGGATAAAGGAGGCTCAAG-----AGATAGAGGGCAGAGCATGA 2237  
 QY 30 -----Lys-----Arg-----Glu-----Arg----- 33  
 DB 2238 GCCAGGTTTAATTTTCTCTGTAGAGATGGTCCCGCAGGAGGTTACTGTGGCTG 2297  
 QY 34 -----Lys-----Phe-----Glu-----AlaAsn-----Cys----- 39  
 DB 2298 GGAGATCTGGGGTATATACACCACCTGAATGATCAGCAGTCAATTCAGAGCTGTGTGGC 2357  
 QY 40 -----GlyLeu----- 41  
 DB 2358 AAAAGGACTGAGACCCAGATTTCTGTCTCTGTGTAGGTGTCTCTGTACCATCTG 2417  
 QY 42 -----Asp-----PheIleIlePhe-----Trp-----Phe-----Trp 50  
 DB 2418 CAGACAGACATCTTC-----ATCTTTTACTATGGCTGTCTCCCTGAATTTACAGCAGTGG 2474  
 QY 51 -----IleLeuLeu----- 54  
 DB 2475 CCAAGCCATTACTG 2534  
 QY 55 Ser-----HisHis----- 60  
 DB 2535 TCTGTGTAGTACACTG 2594

DB 2298 GGAGATCTGGGTATATACACCACCTGAATGATCAGCAGCTCAATTCAGAGCTGTGTGGC 2357  
 QY 40 -----GlyLeu----- 41  
 DB 2358 AAAAGGACTGAGACCCAGATTTCTGTCTCTGTGAGGTCTCTGTCTACCATCTG 2417  
 QY 42 -----Asp-----PheIleIlePhe-----Trp-----Phe-----Trp 50  
 DB 2418 CAGACAGACATCTTC-----ATCTTTTACTATGGCTGTCTCCCTGAATTTACAGCAGTGG 2474  
 QY 51 -----IleLeuLeu----- 54  
 DB 2475 CCAAGCCATTACTG 2534  
 QY 55 Ser-----HisHis----- 60  
 DB 2535 TCTGTGTAGTACACTG 2594  
 QY 61 Glu-----Ser-----LeuLeu-----Cys----- 65  
 DB 2595 GACTTATGTCTGTCTG 2654  
 QY 66 -----ProPro-----Ser-----Pro-----Lys 70  
 DB 2655 CTGACACCTCGGTGTCAGCTGTGTGACCTCCGACGAGCAGTTCCTCAGGGGACTAAAA 2714  
 QY 71 -----GluVal-----Thr-----Cys-----ArgGluMetLeuThr 79  
 DB 2715 TAATGACTAGTCAATTCAGAGTCCCTCATGCTGAATGTTAAACCAAGG----- 2762  
 QY 80 GlyGlyCysLeuProTrp-----AlaThr----- 87  
 DB 2763 -----TGC-----CCCTGGGTGATAGTTTGTAGTCTCTGCACTCTGGTTGGAAGGAAGT 2813  
 QY 88 -----Arg-----Ser-----HisLeu-----Gly-----ArgArg 94  
 DB 2814 GACTACGGAAGCCATCTGCTCCCTGCTGGGAGCTTCCACCTCATGCCAGTGTTCAGAGA 2873  
 QY 95 LysCys-----Ser 97  
 DB 2874 TCTGTGGGAGC 2885

RESULT 3  
 US-08-110-683-3  
 Sequence 3, Application US/08110683  
 Patent No. 5789192  
 GENERAL INFORMATION:  
 APPLICANT: Moore, Kevin W.  
 APPLICANT: Liu, Ying  
 APPLICANT: Ho, Alice Suk-Yue  
 APPLICANT: Hsu, Di-Hwei  
 APPLICANT: Bazan, J. Fernando  
 APPLICANT: Tan, Jimmy C.  
 APPLICANT: Chou, Chuan-Chu  
 TITLE OF INVENTION: Mammalian Receptors for Interleukin-10  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: DNAX Research Institute  
 STREET: 901 California Avenue  
 CITY: Palo Alto  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94304-1104  
 COMPUTER READABLE FORM: disk  
 MEDIUM TYPE: Floppy  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/110,683  
 FILING DATE: 23-AUG-1993





us-09-854-133-586.rni

Tue May 13 12:12:50 2003

```

Db 2814 GGACTACGGAGCAATCTGCCCCCTGGGGAGCTTCCACCTCATGCGCAGTGTTCAGAGA 2873
QY 95 LysCys---Ser 97
Db 2874 TCTTGTGGGAGC 2885

RESULT 5
US-08-472-097-3
; Sequence 3, Application US/08472097
; Patent No. 5985828
; GENERAL INFORMATION:
; APPLICANT: Moore, Kevin W.
; APPLICANT: Liu, Ying
; APPLICANT: Ho, Alice Suk-Yue
; APPLICANT: Hsu, Di-Hwei
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Tan, Jimmy C.
; APPLICANT: Chou, Chuan-Chu
; TITLE OF INVENTION: Mammalian Receptors for Interleukin-10
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,097
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,683
; FILING DATE: 23-AUG-1993
; APPLICATION NUMBER: US 08/011,066
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Chang, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0335K1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3520 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 80..1807
; US-08-472-097-3

Alignment Scores:
Pred. No.: 4.87e+04 Length: 3520
Score: 283.00 Matches: 63
Percent Similarity: 25.66% Conservative: 15
Best Local Similarity: 20.72% Mismatches: 10
Query Match: 53.20% Indels: 216
DB: 2 Gaps: 62

US-09-854-133-586 (1-97) x US-08-472-097-3 (1-3520)
QY 1 Glu-----ValGluValSerArg-----HisAla 9

```

RESULT 6  
 US-09-439-672-3  
 ; Sequence 3, Application US/09439672  
 ; Patent No. 6423500  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Moore, Kevin W.  
 ; APPLICANT: Liu, Ying  
 ; APPLICANT: Ho, Alice Suk-Yue  
 ; APPLICANT: Hsu, Di-Hwei  
 ; APPLICANT: Bazan, J. Fernando

APPLICANT: Tan, Jimmy C.  
 APPLICANT: Chou, Chuan-Chu  
 TITLE OF INVENTION: Mammalian Receptors for Interleukin-10  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: DNAX Research Institute  
 STREET: 901 California Avenue  
 CITY: Palo Alto  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94304-1104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/439,672  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/110,683  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ching, Edwin P.  
 REGISTRATION NUMBER: 34,090  
 REFERENCE/DOCKET NUMBER: DX0335K1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-852-9196  
 TELEFAX: 415-496-1200  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3520 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 80..1807  
 US-09-439-672-3

## Alignment Scores:

Pred. No.: 4,87e-04 Length: 3520  
 Score: 283.00 Matches: 63  
 Percent Similarity: 25.66% Conservative: 15  
 Best Local Similarity: 20.72% Mismatches: 10  
 Query Match: 53.20% Indels: 216  
 DB: 4 Gaps: 62

US-09-854-133-586 (1-97) x US-09-439-672-3 (1-3520)

Qy 1 Glu-----ValGluValSerArg-----Asp-----HisAla 9  
 Db 2001 GAGTGAGGCTCTGTAGATACAGACAGAGCTGAGCAGGATGACAGACAGCTCTCTATGCC 2060  
 Qy 10 Ser-----Leu-----Gly-----12  
 Db 2061 TCAGGGCTGCTCTACACTGGAAGGACCTGTGTTGGGTAACTCAGGCTTCTCGG 2120  
 Qy 13 -----Asp-----SerGlu-----Thr-----LeuSer-----Glu 21  
 Db 2121 ATGGGTAAAGACTAGTCTGAAGTCAGCTGAGCCTGAGCCTGGATGCTGCGGAGGTGTGGAG 2180  
 Qy 22 -----Leu-----Arg-----Lys-----LysGluArg-----Lys-----29  
 Db 2181 TGCTAGCCTGCTACAGGATAAGGGAAGGCTCAAG---AGATAGAAGGGCAGACATGA 2237  
 Qy 30 -----Lys-----Arg-----Glu-----33  
 Db 2238 GCACAGTTTAAATTTCTCTGTAGATGTTCCCGCAGGATGTTACTTGTGCTG 2297

Qy 34 ---Lys-----Phe-----Gln-----AlaAsn-----Cys--- 39  
 Db 2298 GGAGATCTTGGGTATACACACCCCTGAATGATCAGCCAGTCAATTCAGAGCTGTGGC 2357  
 Qy 40 -----GlyIle-----41  
 Db 2358 ARAAGGGACTGACACCCAGAAATTTCTGTCTCTGTGAGGTGCTCTGTCTACCATCTG 2417  
 Qy 42 -----Asp-----PheIleIlePhe-----Trp-----Phe-----Trp 50  
 Db 2418 CAGACAGACATCTTC---ATCTTTTACTATGCTGTGCTCCCTGAATTAACGACAGTGG 2474  
 Qy 51 -----IleLeuLeu-----Phe 54  
 Db 2475 CCAAGCCATTACTCTCTGCTGCTCACTGTTGTGACGTGACACGACGACGACGCTGTCTG 2534  
 Qy 55 Ser-----HisHis-----Trp-----Ile-----Gln-----60  
 Db 2535 TCTGTGTAGTACACTACCTTTAGGTGGCCTTTGGGCTTGAGCTGGCCAGGCTTAG 2594  
 Qy 61 Glu---Ser-----LeuLeu-----Cys-----65  
 Db 2595 GACTATGCTGCTGCTTTTCTGCTCTAATCTTAATCTGACAGCCAGACAGAGGCTGCTGG 2654  
 Qy 66 -----ProPro-----Ser-----Pro-----Lys 70  
 Db 2655 CTGACACCTCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2714  
 Qy 71 -----GluVal-----Thr-----Cys-----ArgGluMetLeuThr 79  
 Db 2715 TAATGACTAGTCAATTCAGAAAGTCCCTCATCTGAATGTTTAACCAAGG-----2762  
 Qy 80 GlyGlyCysLeuProTrp-----AlaThr-----87  
 Db 2763 -----TGC-----CCCTGGGGTGATGTTTAGTCTCTGACAGCTTGGGTTGAAAGAAAGT 2813  
 Qy 88 -----Arg-----Ser-----HisLeu-----Gly-----ArgArg 94  
 Db 2814 GGACTACGGAAGCCATCTGTCCCTGGGAGCTTCCACCTCATGCCAGTGTTCAGAGA 2873  
 Qy 95 LysCys---Ser 97  
 Db 2874 TCTTGTGGGAGC 2885

RESULT 7  
 PCT-US93-11638-3  
 ; Sequence 3, Application PC/TUS9311638  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: Mammalian Receptors For Interleukin-10 (IL-10)  
 ; NUMBER OF SEQUENCES: 12  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: Apple Macintosh  
 ; OPERATING SYSTEM: Macintosh 6.0.8  
 ; SOFTWARE: Microsoft Word 5.1a  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/11638  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/110,683  
 ; FILING DATE: 23-AUG-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/011,066  
 ; FILING DATE: 29-JAN-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/989,792  
 ; FILING DATE: 10-DEC-1992  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3520 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single

us-09-854-133-586.rni

Tue May 13 12:12:50 2003

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
PCT-US93-11638-3

Alignment Scores:  
Pred. No.: 4.87e+04  
Score: 283.00  
Percent Similarity: 25.66%  
Best Local Similarity: 20.72%  
Query Match: 53.20%  
DB: 5  
Gaps: 62  
Indels: 216  
Mismatch: 10  
Conservative: 15  
Matches: 63  
Length: 3520

US-09-854-133-586 (1-97) x PCT-US93-11638-3 (1-3520)

QY	1	Glu	-----ValGluValSerArg-----Asp-----HisAla 9
DB	2001	GAGTGGGCTGTGTAGATACACGAGAGCTGAGCAGGAGTTGACAGAGACCTCTCATGCC 2060	
QY	10	Ser	-----Leu-----Gly-----12
DB	2061	TCAGGGCTGGCTTCACACTGGAAGACCTGTGTTGGTGTAACTCAGGCTTTCTGG 2120	
QY	13	Asp	-----SerGlu-----Thr-----LeuSer-----GlnThr-----Glu 21
DB	2121	ATGTGTAGACTGTAGTCTGAAGTCAGCTGAGCTGAGCTGTGATGTCTCGGAGGTGTGGAG 2180	
QY	22	Leu	-----Arg-----Lys-----LysGluArg-----Lys-----Lys-----29
DB	2181	TGGCTAGCCTGCTACAGGATAAGGAAGGCTCAAG-----AGATAGAGGGCAGAGCATGA 2237	
QY	30	Lys	-----Arg-----Glu-----Arg-----33
DB	2238	GCAGGTTTAAATTTGCTCTGTAGAGATGTCCTCCAGCAGGATGGTGTACTTGTGGCTG 2297	
QY	34	Lys	-----Phe-----Gln-----AlaAsn-----Cys-----39
DB	2298	GGAGATCTGGGGTATACACACCTGGAATGATCAGCAGCTCAATTCAGAGCTGTGGC 2357	
QY	40	-----	-----GlyIle-----41
DB	2358	AAAGGAGCTGAGACCAAGATTTCTGTCTCTGTGTGAGGTGTCTGTCTACCCATCTG 2417	
QY	42	Asp	-----PheIleIlePhe-----Trp-----Ile-----Phe-----Trp 50
DB	2418	CAGACAGACATCTTC-----ATCTTTTACTATGGCTGTGTCCCTGGAATACAGCAGTGG 2474	
QY	51	LeuLeuLeu	-----IleLeuLeu-----Phe 54
DB	2475	CAAGCCATTACTCCCTGCTGCTACTGTGTGAGCTGACAGCAGACAGAGCTGTCTG 2534	
QY	55	Ser	-----HisHis-----Trp-----Ile-----Gln-----60
DB	2535	TCTGTGTAGTACACTACCTTTAGTGGCTTTAGTGGCTTTAGTGGCTTTAGTGGCTTTAG 2594	
QY	61	Glu	-----Ser-----LeuLeu-----Cys-----65
DB	2595	GACTTATGTCTGCTTTGCTGTCTATCTCTACTGACAGCCAGACAGAGGCTGTGGG 2654	
QY	66	ProPro	-----Pro-----Lys 70
DB	2655	CTGACACCTCCGTTGTTGCTGTGTGACCTCCGACAGCAGAGCTTCTCAGGGGACATARA 2714	
QY	71	-----	-----GluVal-----Thr-----Cys-----ArgGluMetLeuThr 79
DB	2715	TAATGACTAGTCAATTCAGAGTCCCTCATGCTGTAATGTTAACCAAG-----2762	
QY	80	GlyGlyLeuProTrp	-----AlaThr-----87
DB	2763	-----TGC-----CCCTGGGTGATGTTAGTCTCTCAACCTCTGGTGTGGAGGAAGT 2813	
QY	88	Arg	-----Ser-----HisLeu-----Gly-----ArgArg 94
DB	2814	GGACTACGGAAGCAATCTGTCCCTCCCTGGGAGGCTTCCACCTCATGCTCCAGTGTTCAGAGA 2873	

QY 95 LysCys---Ser 97  
DB 2874 TCTTGTGGGAGC 2885

RESULT 8  
US-08-444-734A-1  
Sequence 1, Application US/08444734A  
Patent No. 5610282  
GENERAL INFORMATION:  
APPLICANT: Sibley, David R.  
APPLICANT: Monsma, Frederick J.  
APPLICANT: Mahan, Lawrence C.  
APPLICANT: McVittie, Loris D.  
TITLE OF INVENTION: receptor linked to adenylyl cyclase activation and expression of the receptor protein in plasmid-transfected cells  
TITLE OF INVENTION: cell lines  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,734A  
FILING DATE:  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/029,917  
FILING DATE: 03-MAR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/548,714  
FILING DATE: 06-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: NIH065.001FW1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3025 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: PB73D1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 293..1756  
US-08-444-734A-1

Alignment Scores:  
Pred. No.: 2.88e+04  
Score: 282.50  
Percent Similarity: 29.73%  
Best Local Similarity: 22.01%  
Query Match: 53.10%  
DB: 1  
Length: 3025  
Matches: 57  
Conservative: 20  
Mismatch: 11  
Indels: 171  
Gaps: 62

US-09-854-133-586 (1-97) x US-08-444-734A-1 (1-3025)

; PRIOR APPLICATION NUMBER: US 60/119,917		
; PRIOR FILING DATE: 1999-02-12		
; NUMBER OF SEQ ID NOS: 1304		
; SOFTWARE: Patent.pm		
; SEQ ID NO 319		
; LENGTH: 1001		
; TYPE: DNA		
; ORGANISM: Homo Sapiens		
; FEATURE:		
; NAME/KEY: allele		
; LOCATION: 501		
; OTHER INFORMATION: 12-374-261 : polymorphic base G or A		
; NAME/KEY: misc.binding		
; LOCATION: 502..521		
; OTHER INFORMATION: 12-374-261.mis1, potential complement		
; NAME/KEY: misc.binding		
; LOCATION: 482..500		
; OTHER INFORMATION: 12-374-261.mis2		
; NAME/KEY: primer_bind		
; LOCATION: 741..761		
; OTHER INFORMATION: upstream amplification primer, complement		
; NAME/KEY: primer_bind		
; LOCATION: 262..280		
; OTHER INFORMATION: downstream amplification primer		
; NAME/KEY: misc.binding		
; LOCATION: 489..513		
; OTHER INFORMATION: 12-374-261 potential probe		
US-09-641-638-319		
Alignment Scores:		
Pred. No.:	449	Length:
Score:	282.00	Matches:
Percent Similarity:	33.93%	Conservative:
Best Local Similarity:	26.79%	Mismatches:
Query Match:	53.01%	Indels:
DB:	4	Gaps:
		45
US-09-854-133-586 (1-97) x US-09-641-638-319 (1-1001)		
QY	1	GluValGlu-----ValSerArgAspHisAla--SerLeu-----Gly---AspSer 14
Db	128	GAGACAGAGCTTGCGACTGAGCCAGCATCGGCCACTGCCTCCAGCCTGGGTGACAGGC 18
QY	15	GluThrLeuSerGluThrGluLeuArgLys---LysGluArgLys-----28
Db	188	GAGACTCTGCTCAA---AAA---AAGAAAAAGAAAGAAAGAAACATGGGTGCAGTCAC 241
QY	29	LysLysArgGlu-----Arg---LysPheGln-----36
Db	242	CCAAAGAGAGGAGGAGGCGACGGACAGGCAGGAGAAA---CAGCTGTATGACCAGCACCT 298
QY	37	Ala-----Asn-----Cys-----Gly 40
Db	299	CGGTTTCCTCCATTGTGACAAAGAATCCAAAGCTCAAGAGGGGAAGTGTCTTTGTCTGGA 358
QY	41	Ile---AspPheIleIle-----PheTrpIlePheTrpIleLeu---Leuphe 54
Db	359	GCCTCGGAG---CTGGTAGCTTAGGAGCCTTACTGCCCC---TGGGTCTCGGCCCTG---409
QY	55	Ser-----His-----HisTrpIleGlnGluSer 62
Db	410	GATGGAGAGAAGGGGAGGCCCATGTGTGAGGGCACTCTTCAT---GTCCAG---AGC 463
QY	63	LeuLeuCys-----ProProSerProLysGlu-----Val-----72
Db	464	CTAGCTTGCTTTCAGACCT---CCTCAAGAGAGGAGGAGRGTTTGAAGGATGCC 517
QY	73	-----Thr-----Cys-----Arg---75
Db	518	GGNAGCCAAACTGCTTCTCTCCATACTTCTTGTAATTTCCTTAAGCAGGAAAGACCT 577
QY	76	-----Glu---MetLeu---Thr---Gly---Gly---82





Alignment Scores:		
Pred. No.:	3.14e-04	2907
Score:	279.50	72
Percent Similarity:	24.57%	14
		Conservative:

Db 1153 GTCAGCGGGCTGAACCTTCGTTTTCATCTTTTCGGGCTCTCTTTTGGGCTGG 1094

US-09-854-133-586 (1-97) x US-09-018-635-26 (1-2907)

238  
53

-----Gly---AspSer 14  
|||  
|||  
GCCTGGGTGCAGAGC 1739  
  
slysArgGluArg--- 33  
|||  
|||:|||||  
AAAAAGAAAGAAAGACA 1691  
----- 33  
  
ETGTGAAATGAAAGCCT 1631  
  
ASN-----Cys 39  
|||  
|||  
TAATCATCATCAGCTGT 1571  
-----  
|||  
|||  
TTCCCCATATTTCTG 1511  
  
IlePhe---TrpIle 48  
:|||| :|||  
C|TTTTGTTATAT 1451  
-----  
LeuPheSerHis 56  
||||| |||  
ATCCTCTTC--CAC 1394  
----- 57  
  
TTTGGTACTTTGGG 1334  
----- 59  
  
GGTGGCGCTGAAGC 1274  
-----Ser-- 62  
|||  
CTGCTCCTCCACCA 1214  
-----  
ProProSerPr 69  
|||||||  
|  
GAGCCACCCCTCAC 1154  
----- 71  
TTTTTTTGAGGCTCG 1094  
----- 73  
AGTTTGTGAAAGGC 1034  
----- 76  
AGAGTCATCATGCC 974  
-----Thr-- 79  
|||  
GC|TTTCCACAGG 914  
-----  
Thr-----Ar 88  
:|:  
|  
ACTCAGATTTCAA 854  
-----ArgLy 95  
|||  
GGCTCTCAGGAA 794

```

; SEQ ID NO 3
; LENGTH: 2907
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (125)..(1627)
; NAME/KEY: unsure
; LOCATION: 1894
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 1990
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 2388
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 2506
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 2509
; OTHER INFORMATION: unknown
; US-09-457-642-3

```

Length:	2907
Matches:	72
Conservative:	14
Mismatches:	6
Indels:	258
Gaps:	53

US-09-854-133-586 (1-97) x US-09-467-642-3 (1-2907)

OY	1	GluValGlu-----ValSerArgAspHisAla-----SerLeu-----Gly-----AspSer	14
DB	1798	GAGACGGAGGTGCGACTGAGCGCGAGATCACGCCACTGCATCCAGCGTGGGTGCACAGAC	1739
OY	15	GluThrLeuSerGlnThrGluLeuArgLysLysGluArgLysLysLysAsqGluuArg---	33
DB	1738	GAGACTCTGCTCAA-----AAAAAAAAA-----AAAAAGAAAAGAAAGAAAGCA	1691
OY	33	-----	33
DB	1690	GACTATCAGGGGCTATTATTAGGAACCATCTCTCTGTGAAATCTGTGGAAATGAAAGCCT	1631
OY	34	-----Lys-----PheGln-----Ala-----Asn-----Cys	39
DB	1630	GTTTTCAGTTCATGCCAAGTCTTTTCATGTTCGCCAGCGATCTCTTAATCATCATCAGCTGT	1571
OY	40	-----Gly-----Ile-----Ile-----	41
DB	1570	TCGTTTAACAAATGGGTAATTTTAGAAATGGCAGCCCAAGTTTCCTTCCCATATTTCTG	1511
OY	42	-----Asp-----Phe-----IleIlePhe-----TrpIle	48

**Alignment Scores:**



Tue May 13 12:12:50 2003

APPLICATION NUMBER: US/09/041,886

FILE DATE: 05/09/04  
CLASSIFICATION: A61K 31/56  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 2626  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4608 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..4342  
US-09-041-886-24

Alignment Scores: 1.67e+05 Length: 4608  
Pred. No.: 279.50 Matches: 66  
Score: 25.00% Conservativity: 18  
Percent Similarity: 19.64% Mismatches: 9  
Best Local Similarity: 52.54% Indels: 243  
Query Match: 4 Gaps: 71  
DB:

US-09-854-133-586 (1-97) x US-09-041-886-24 (1-4608)

QY 1 Glu---Val---Glu---Val---SerArgasp---His-----Ala----- 9  
DB 2049 GAATAGTACACAGAGGTGTTGGCTC---CAGTGTTCATCTCACCCCTCGGGTCGT 1993  
QY 10 ---Ser---Leu---Glyasp---SerGluThrLeu---Ser--- 18  
DB 1992 CTTTCGTGTGGAATTTATAGCGGTAAATAATCC---ATTTGTGTCTCTGANGAGG 1936  
QY 19 ---Gln---Thr---Glu---LeuArg 23  
DB 1935 AGCAGCAGCAGTAACTTTGATGACTTTGAATTCACCTCCAGGGAGAGCTTCTGAGG 1876  
QY 24 Lys---Lys---Glu---ArgLys---Lys-----LysArgGlu---Arg 33  
DB 1875 CGGGGCACTTGGCAGCTGAGAAGGTAAACCACTGTTATATCATCAGTAGACGCCCGG 1816  
QY 34 ---Lys-----PheGlnAla--- 37  
DB 1815 ACCATAGCGATTATAAGCTAAGAACTATATTCGGTGAATTTTTCAGGCTTC 1756  
QY 38 ---Asn-----Cysgly----- 40  
DB 1755 CAGTTTATAGATAGTCCATCACTCTATATTTCTTTCTTCTGTGTGGACACCTCAT 1696  
QY 41 ---Ile---Asphe---Ile---Ile-----Phe----- 46  
DB 1695 GCAGAACAAATCTGAACCTTGGACTGGACCGTTGATAGCAGGAGGGGTCCAGGTAAT 1636  
QY 47 ---Trp-----IlePhe-----Trp----- 50  
DB 1635 AGAATTGAGGTAGGTAGGTAGATACAGCTTCAGGTTTCTACTGGCCCTGGAACTTG 1576  
QY 51 ---Ile-----Leu-----Phe--- 54  
DB 1575 CAATCAGGCTGTGTGGCCACCTTGATGGGTGTGAGAACTCTCTCCCGTCCCATTCATT 1516  
QY 55 ---SerHis-----His-----TripleGln-----GluSer 62  
DB 1515 GTAAGCCCAACACTGGAAGGTGTACATGGCTTCTGTGCTTCAGGTTTCCACAGTAG 1459  
QY 63 Leu-----Leu-----Cys-----Pro----- 66

DB 1510 CACTCCAGCTTGACCCACTCGCTTCTTACAGTCCACCTCTGCTTTTGTATATTT 1451  
QY 49 ---Phe---Trp---Ile---Leu-----LeuPheSerHis 56  
DB 1450 GGTGTACTGTCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1394  
QY 57 ---His----- 57  
DB 1393 CCAAGTCTCTTCTTCAACCCATTAGAGCTGTTCACACTTGCCTTTGGTACTTTGGG 1334  
QY 58 ---Trp-----Ile----- 59  
DB 1333 ATTCTTCTCCAGGAGGGTGTGTGAGACGGTGGGCTTGGATGGTGGCTGAAGC 1274  
QY 60 ---GlnGlu-----Ser--- 62  
DB 1273 GGCTCTGGGAGGTGAGGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1214  
QY 63 ---Leu---Leu---Cys-----ProProSerPr 69  
DB 1213 GACCAATCTGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1154  
QY 69 Olys-----Glu----- 71  
DB 1153 GTCAGCGGGCTGAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1094  
QY 72 ---Val---Thr----- 73  
DB 1093 TGATCTGGGAGCTTGAGTAGGAGAACACAGATCTTCTGCTCCAGTTTGTCAAAGGC 1034  
QY 74 ---Cys---Arg---Glu----- 76  
DB 1033 TGCCTCAGATCTGTGCACACAGAGCTTGTAAAGCTGCTTTCAGAGTCAATCTCC 974  
QY 77 -Met-----Leu-----Thr--- 79  
DB 973 AATGGTGGTGGAGGATCCGTAGTGCCTTGGGGTTCCTGCTGGTGGCTTTTCCACAGG 914  
QY 80 ---Gly---GlyCysLeu---Pro---Trp-----Ala---Thr-----Ar 88  
DB 913 CCTGTGCTGGCTGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 854  
QY 88 gSerHisLeu-----Gly-----Arg-----ArgLy 95  
DB 853 AGCCTTTTGGCCATCGTAGGAGGTAGGCTCGGGCTCATCCAGGTGGCTCTCCAGGAA 794  
QY 95 s-----Cys-----Ser 97  
DB 793 GCGCAGCATCTTCTGCTGGAAGGTCTCA 766

RESULT 15  
US-09-041-886-24/c  
Sequence 24, Application US/09041886  
Patent No. 6235872

GENERAL INFORMATION:  
APPLICANT: Bredesen, Dale E.  
APPLICANT: Rabizadeh, Sharrow  
TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
TITLE OF INVENTION: Polypeptides and Methods of Use  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

```
Db 1458 CTGAAGGACCCAGGCTGTGTTGTTATTCATGCTGCTCCCTGTTGTACCTTCTCTGGA 1399
QY 67 -----Pro-----SerPro----- 69
Db 1398 GAAAGAGACCGTGAAGTTTGAATGTTCCCTTTCCCTTCTGAGGTGGCGCCAGCTGAG 1339
QY 70 -----Lys-----Glu-----Val----- 72
Db 1338 ACGGACAAATCGGCTGGAAACCAAGACAGGACCAATCTCTGGGAGCCGAAGGAGGAC 1279
QY 73 Thr-----Cys-----Arg-----GluMetLeu---ThrGly---Gly--- 81
Db 1278 ACTGGAGCTTGGGATTGAGGCTTAGGACAATGAG--CTGTCCACTGGTCTGGGCATT 1222
QY 82 -----Cys-----Leu-----Pro--- 84
Db 1221 TCCAGCCTCATTTTCAGCCACACATTCATGATAAAGCCTTCATCTGACTTCACCCCAAG 1162
QY 85 Trp-----Ala-----ThrArg-----Ser 89
Db 1161 TATCCGTAAGTTGCTTCTCCCACTATCTGAAATAATCACTAGGAAATGACCACATCTCC 1102
QY 90 -----His---Leu-----Gly---Arg-----ArgLysCysSer 97
Db 1101 ATTCTTCATCCAATTCACAGTGGGCACAGGCTTCCAGAGACTGTACA 1054
```

Search completed: May 11, 2003, 19:33:35  
Job time : 49 secs

Tue May 13 12:12:50 2003

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 19:06:36 ; Search time 90 Seconds  
(without alignments)  
1339.360 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHSAGLSDSTLST.....LFGGCLPWATSHLGRKCS 97

Scoring table: BLOSUM62  
Xgapop 0.0, Xgapext 0.5  
Ygapop 0.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1556566

Minimum DB seq length: 0

Maximum DB seq length: 5000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODEL=frame+g2n.model -DEV=xlh

-Q=/cgn2.1/USPTO\_spool/US09854133/runat\_05052003.174428.1338/app\_query.fasta\_1.263

-DB=PublishedApplications\_NA -QMT=fastap -SUFFIX=rnpb -MINMATCH=0.1

-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100

-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=5000 -USER=US09854133@cgn1.1.91@runat\_05052003.174428.1338 -NCPUN=6

-ICPU=3 -NO\_ALPAY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG

-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=0 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_NA:

- 1: /cgn2.6/ptodata/2/pubpna/US07\_PUBCOMB.seq:
- 2: /cgn2.6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:
- 3: /cgn2.6/ptodata/2/pubpna/US06\_PUBCOMB.seq:
- 4: /cgn2.6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:
- 5: /cgn2.6/ptodata/2/pubpna/US07\_PUBCOMB.seq:
- 6: /cgn2.6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:
- 7: /cgn2.6/ptodata/2/pubpna/US08\_PUBCOMB.seq:
- 8: /cgn2.6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:
- 9: /cgn2.6/ptodata/2/pubpna/US09\_PUBCOMB.seq:
- 10: /cgn2.6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:
- 11: /cgn2.6/ptodata/2/pubpna/US10\_PUBCOMB.seq:
- 12: /cgn2.6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:
- 13: /cgn2.6/ptodata/2/pubpna/US60\_PUBCOMB.seq:
- 14: /cgn2.6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	532	100.0	337	9	US-09-854-133-442
2	532	100.0	337	10	US-09-738-973-442
3	523	98.3	2239	9	US-09-854-133-440
4	523	98.3	2239	10	US-09-738-973-440

5	339.5	63.8	1861	9	US-10-163-866-30
6	339.5	63.8	2482	9	US-10-163-866-29
7	339	63.7	572	9	US-10-046-935-1307
8	339	63.7	572	9	US-09-878-178-1307
9	339	63.7	572	9	US-10-146-502-1307
10	298.5	56.1	1687	9	US-09-764-891-6421
11	298.5	56.1	1688	9	US-09-764-891-6422
12	296.5	55.7	2000	9	US-10-163-866-34
13	292.5	55.0	3995	10	US-09-919-497-9
14	287.5	54.0	2427	10	US-09-254-783A-2
15	287.5	54.0	2427	12	US-10-152-058-2
16	284.5	53.5	3243	9	US-09-764-847-1176
17	284.5	53.5	3243	10	US-09-764-847-1176
18	283.5	53.3	3162	10	US-09-764-877-2937
19	281.5	52.9	2203	10	US-10-224-562-1
20	281.5	52.9	2203	10	US-09-801-861-1
21	281.5	52.9	3308	9	US-10-147-026-3
22	281.5	52.9	4344	10	US-09-880-107-1754
23	281.5	52.8	4168	12	US-10-042-417-27
24	279.5	52.5	2907	10	US-09-912-962-26
25	279.5	52.5	4870	12	US-10-044-090-121
26	279	52.4	3351	9	US-10-092-154-1138
27	279	52.4	3351	10	US-09-764-847-1138
28	278.5	52.3	2281	9	US-10-071-766-71
29	278.5	52.3	2395	9	US-10-152-661-590
30	278.5	52.3	2395	9	US-09-866-050A-590
31	278	52.3	2495	9	US-10-091-504-1255
32	278	52.3	2495	10	US-09-764-869-1255
33	277.5	52.2	2507	10	US-09-728-628-9
34	277.5	52.2	2808	10	US-09-925-301-397
35	277.5	52.2	2808	10	US-09-880-107-3380
36	277	52.1	2969	9	US-09-954-531-179
37	276.5	52.0	1611	9	US-10-124-986-2
38	276.5	52.0	1611	10	US-09-981-649A-2
39	276.5	52.0	2569	9	US-10-224-951-1
40	276.5	52.0	2872	10	US-09-764-877-3199
41	276.5	52.0	2872	10	US-09-764-877-3196
42	276.5	52.0	2903	9	US-09-984-827-138
43	276.5	52.0	4152	10	US-09-925-300-359
44	276.5	52.0	4766	9	US-09-764-868-1443
45	276	51.9	2018	9	US-10-091-504-1568

## ALIGNMENTS

RESULT 1  
US-09-854-133-442  
Sequence 442, Application US/09854133  
Publication No. US20020183499A1  
GENERAL INFORMATION:  
APPLICANT: Lodes, Michael J.  
APPLICANT: Mohamath, Raodoh  
APPLICANT: Henderson, Robert A.  
APPLICANT: Benson, Darin R.  
APPLICANT: Secrist, Heather  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.475C10  
CURRENT APPLICATION NUMBER: US/09/854.133  
CURRENT FILING DATE: 2001-05-11  
NUMBER OF SEQ ID NOS: 735  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 442  
LENGTH: 337  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-854-133-442

Alignment Scores:  
Pred. No.: 1.77e-06  
Score: 532.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Length: 337  
Matches: 97  
Conservative: 0  
Mismatch: 0

---cccccnnnrrrrggc 124

Db 302 CCCTTTTCAGGAAGAGAGCGCTTTTCAGGAAGAGAGAAAGTGCAGC 348  
 ArgLysCysSer 97  
 |||||

RESULT 4  
US-09-738-973-440  
; Sequence 440, Application US/09738973  
; Patent No. US20020110563A1

us-09-854-133-586.rnpb

Tue May 13 12:12:50 2003

GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indrias, Carol Yoseph  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Elliot, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14  
 ; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FASTSEQ for Windows Version 3.0  
 ; SEQ ID NO 440  
 ; LENGTH: 2239  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-738-973-440

Alignment Scores:  
 Pred. No.: 0.0022 Length: 2239  
 Score: 523.00 Matches: 97  
 Percent Similarity: 84.35% Conservativeness: 0  
 Best Local Similarity: 84.35% Mismatches: 0  
 Query Match: 98.31% Indels: 18  
 DB: 10 Gaps: 1

US-09-854-133-586 (1-97) x US-09-738-973-440 (1-2239)  
 QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 Db 2 GAGTTGAGTGCAGCAGAGATCATGCCAGCTGGTGGTGACGTGACGTGACGTCTGTCTCAACA 61  
 QY 21 GluLeuArgGlyLysGluArgLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
 Db 62 GAATTAAGGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATCCAGGCCAATGTGGC 121  
 QY 41 IleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
 Db 122 ATAGATTTATCATATCTGGATTCTTTGGATTCTTTGTTCTTCATCATCTGGATTCTAG 181  
 QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 Db 182 GAAAGCCTGTGTGTCACCATCTCCAAAGGAGGTACCTGCAGGGAATGTTAAACGGGA 241  
 QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg 93  
 Db 242 GCCTGCCTTCCTGGCAACAAGAGAGCCACCTGGCAGGAGACGCTTTTCAGGAAGACG 301  
 QY 94 -----ArgLysCysSer 97  
 Db 302 CCTTTTCAGGAAGACGCCCTTTTCAGGAAGAGAAAGTGCAGC 346

RESULT 5  
 US-10-163-866-30  
 ; Sequence 30, Application US/10163866  
 ; Publication No. US20030027188A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: EXELIXIS, INC.  
 ; TITLE OF INVENTION: SLC7s AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
 ; FILE REFERENCE: EX02-080C  
 ; CURRENT APPLICATION NUMBER: US/10/163,866  
 ; CURRENT FILING DATE: 2002-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/296,076  
 ; PRIOR FILING DATE: 2001-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/328,605  
 ; PRIOR FILING DATE: 2002-02-15  
 ; PRIOR APPLICATION NUMBER: US 60/357,600  
 ; PRIOR FILING DATE: 2002-02-15

US-09-854-133-586 (1-97) x US-10-163-866-30 (1-1861)  
 QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 11  
 Db 6 GAGTGGAGAAATTCAGACGACGATGCATACACAGGAGTCTTCAGTAGTAATTA 62  
 QY 12 GlyAspSerGluThrLeuSerGlnThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 22  
 Db 63 ---GATCGCTGTGAAGAAACACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 116  
 QY 23 ArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 33  
 Db 117 CGTGAGAGACAGCTGTAAGACGAGAGAGATCATCATACACCAACAGACACACC 176  
 QY 34 LysPheGlnAlaSerCysGlyIleAspPheIleIlePheTrpIlePheTrpIlePheTrp 49  
 Db 177 AAA-----GTGGAAGTTT---GTTTTCTTCCCTCTGTTTATTT 215  
 QY 50 TrpIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 66  
 Db 216 TTCCCCCTGTGTCCCTACTA-----TGG-----TCAGAAAGCCTGTGTGTCCA 260  
 QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyCysLeuProTrpAla 86  
 Db 261 CCATCTCCAAAGGAGGTACCTGCAGGGAATGTTAAACGGAGGCTGCTTCCCTGGGCA 320  
 QY 87 ThrArgSerHisLeuGlyArgArgLysCysSer 97  
 Db 321 ACAAGGAGCCACCTGGCAGGAGAAAGTGCAGC 353

RESULT 6  
 US-10-163-866-29  
 ; Sequence 29, Application US/10163866  
 ; Publication No. US20030027188A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: EXELIXIS, INC.  
 ; TITLE OF INVENTION: SLC7s AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
 ; FILE REFERENCE: EX02-080C  
 ; CURRENT APPLICATION NUMBER: US/10/163,866  
 ; CURRENT FILING DATE: 2002-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/296,076  
 ; PRIOR FILING DATE: 2001-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/328,605  
 ; PRIOR FILING DATE: 2001-10-10  
 ; PRIOR APPLICATION NUMBER: US 60/338,733  
 ; PRIOR FILING DATE: 2001-10-22  
 ; PRIOR APPLICATION NUMBER: US 60/357,253  
 ; PRIOR FILING DATE: 2002-02-15  
 ; PRIOR APPLICATION NUMBER: US 60/357,600  
 ; PRIOR FILING DATE: 2002-02-15

GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indrias, Carol Yoseph  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Elliot, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14  
 ; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FASTSEQ for Windows Version 3.0  
 ; SEQ ID NO 440  
 ; LENGTH: 2239  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-738-973-440

Alignment Scores:  
 Pred. No.: 0.0022 Length: 2239  
 Score: 523.00 Matches: 97  
 Percent Similarity: 84.35% Conservativeness: 0  
 Best Local Similarity: 84.35% Mismatches: 0  
 Query Match: 98.31% Indels: 18  
 DB: 10 Gaps: 1

US-09-854-133-586 (1-97) x US-09-738-973-440 (1-2239)  
 QY 1 GluValGluValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeuSerGlnThr 20  
 Db 2 GAGTTGAGTGCAGCAGAGATCATGCCAGCTGGTGGTGACGTGACGTGACGTCTGTCTCAACA 61  
 QY 21 GluLeuArgGlyLysGluArgLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
 Db 62 GAATTAAGGAAAAAGAAAGAAAGAAAGAAAGAAAGAAATCCAGGCCAATGTGGC 121  
 QY 41 IleAspPheIlePheTrpIlePheTrpIleLeuLeuPheSerHisTrpIleGln 60  
 Db 122 ATAGATTTATCATATCTGGATTCTTTGGATTCTTTGTTCTTCATCATCTGGATTCTAG 181  
 QY 61 GluSerLeuLeuCysProSerProLysGluValThrCysArgGluMetLeuThrGly 80  
 Db 182 GAAAGCCTGTGTGTCACCATCTCCAAAGGAGGTACCTGCAGGGAATGTTAAACGGGA 241  
 QY 81 GlyCysLeuProTrpAlaThrArgSerHisLeuGlyArg 93  
 Db 242 GCCTGCCTTCCTGGCAACAAGAGAGCCACCTGGCAGGAGACGCTTTTCAGGAAGACG 301  
 QY 94 -----ArgLysCysSer 97  
 Db 302 CCTTTTCAGGAAGACGCCCTTTTCAGGAAGAGAAAGTGCAGC 346

RESULT 5  
 US-10-163-866-30  
 ; Sequence 30, Application US/10163866  
 ; Publication No. US20030027188A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: EXELIXIS, INC.  
 ; TITLE OF INVENTION: SLC7s AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
 ; FILE REFERENCE: EX02-080C  
 ; CURRENT APPLICATION NUMBER: US/10/163,866  
 ; CURRENT FILING DATE: 2002-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/296,076  
 ; PRIOR FILING DATE: 2001-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/328,605  
 ; PRIOR FILING DATE: 2002-02-15  
 ; PRIOR APPLICATION NUMBER: US 60/357,600  
 ; PRIOR FILING DATE: 2002-02-15

; NUMBER OF SEQ ID NOS: 54  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 29  
 ; LENGTH: 2482  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-163-866-29

## Alignment Scores:

Pred. No.: 954  
 Score: 339.50  
 Percent Similarity: 57.25%  
 Best Local Similarity: 47.33%  
 Query Match: 63.82%  
 DB: 9  
 Length: 2482  
 Matches: 62  
 Conservative: 13  
 Indels: 49  
 Gaps: 28

US-09-854-133-586 (1-97) x US-10-163-866-29 (1-2482)

QY 1 GluValGluValSerArg---AspHis---Ala---Ser-----Leu 11  
 DB 2 GAG---GAGTGAGAAATGAGAGCAGCATACACAGGTGTTCTGAGTAGTAATTA 58  
 QY 12 GlyAsp---SerGlu---Thr---LeuSer---Gln---ThrGluLeu--- 22  
 DB 59 ---GATCGCTGTGAAGGAAAGAACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 112  
 QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
 DB 113 CGCTGAGAGACAGTCTGAAGCAGAGAGACATCGATCAGTAACACCAAGAGACACC 172  
 QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49  
 DB 173 AAA-----GTTGAAAGTTT---GTTTCTTCCCTCGTTTATT 211  
 QY 50 Trp-----Ile---LeuLeuPheSerHisTrpIleGlnGluSerLeuLeuCysPro 66  
 DB 212 TTCCCGCGTGTCCCTACTA-----TGG---TCAGAAAGCCTGTGTGTCCA 256  
 QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
 DB 257 CCATCTCCAAAGAGGTACCTGCAGGAAATGTTAACGGGAGGCTGCTTCCCTGGGCA 316  
 QY 87 ThrArgSerHisLeuGlyArgLysCysSer 97  
 DB 317 ACAAGAGCCACCTGGGAGGAGAAAGTGCAGC 349

## RESULT 7

US-10-046-935-1307  
 ; Sequence 1307, Application US/10046935  
 ; Patent No. US20020156011A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jiang, Yuqiu  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Wang, Aljun  
 ; APPLICANT: Stolk, John A.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; FILE REFERENCE: 210121.527C1  
 ; CURRENT APPLICATION NUMBER: US/10/046.935  
 ; CURRENT FILING DATE: 2002-01-15  
 ; NUMBER OF SEQ ID NOS: 2239  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1307  
 ; LENGTH: 572  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: 9, 19, 461, 497, 500, 502  
 ; OTHER INFORMATION: n = A,T,C or G

US-10-046-935-1307

## Alignment Scores:

Pred. No.: 6.37  
 Score: 339.00  
 Percent Similarity: 57.25%  
 Best Local Similarity: 47.33%  
 Query Match: 63.72%  
 DB: 9  
 Length: 572  
 Matches: 62  
 Conservative: 13  
 Indels: 50  
 Gaps: 28

US-09-854-133-586 (1-97) x US-10-046-935-1307 (1-572)

QY 1 GluValGluValSerArg---AspHis---Ala---Ser-----Leu 11  
 DB 22 GAGGTGGAG-----AATTGAGAGCAGCATACACAGGTGTTCTGAGTAGTAATTA 75  
 QY 12 GlyAsp---SerGlu---Thr---LeuSer---Gln---ThrGluLeu--- 22  
 DB 76 ---GATCGCTGTGAAGGAAAGAACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 129  
 QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
 DB 130 CGCTGAGAGACAGTCTGAAGCAGAGAGACATCGATCAGTAACACCAAGAGACACC 189  
 QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49  
 DB 190 AAA-----GTTGAAAGTTT---GTTTCTTCCCTCGTTTATT 228  
 QY 50 Trp-----Ile---LeuLeuPheSerHisTrpIleGlnGluSerLeuLeuCysPro 66  
 DB 229 TTCCCGCGTGTCCCTACTA-----TGG---TCAGAAAGCCTGTGTGTCCA 273  
 QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
 DB 274 CCATCTCCAAAGAGGTACCTGCAGGAAATGTTAACGGGAGGCTGCTTCCCTGGGCA 333  
 QY 87 ThrArgSerHisLeuGlyArgLysCysSer 97  
 DB 334 ACAAGAGCCACCTGGGAGGAGAAAGTGCAGC 366

## RESULT 8

US-09-878-178-1307  
 ; Sequence 1307, Application US/09878178  
 ; Patent No. US20020177552A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jiang, Yuqiu  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Secrist, Heather  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; FILE REFERENCE: 210121.527  
 ; CURRENT APPLICATION NUMBER: US/09/878,178  
 ; CURRENT FILING DATE: 2001-06-08  
 ; NUMBER OF SEQ ID NOS: 2237  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1307  
 ; LENGTH: 572  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: (1),---(572)  
 ; OTHER INFORMATION: n = A,T,C or G

## Alignment Scores:

Pred. No.: 6.37  
 Score: 339.00  
 Percent Similarity: 57.25%  
 Best Local Similarity: 47.33%  
 Query Match: 63.72%  
 DB: 9  
 Length: 572  
 Matches: 62  
 Conservative: 13  
 Indels: 50  
 Gaps: 28

US-09-854-133-586 (1-97) x US-09-878-178-1307 (1-572)

us-09-854-133-586.rnpb

Tue May 13 12:12:50 2003

QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
 Db 22 GAGGTGGAG-----AATTGAGAGCAGATGATACACAGAGTGTCTGAGTAGTAATTA 75  
 QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
 Db 76 ---GATGCTGTGAAGGAAAGACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 129  
 QY 23 Arg-----LysLys---Glu---Arg---Lys-----LysLys-----ArgGluArg 33  
 Db 130 CGCTGAGAGACAGCTGTAAGAGCAGAGAGATCATCATCAGTACAGTAACACCAAGACACACC 189  
 QY 34 LysPheGlnAlaAsnCysGlyLeuAsp---PheIleIlePheTrp-----Ile---Phe 49  
 Db 190 AAA-----GTTGAAAGTTT---GTTTCTTCCCTCTCTCTTTTATTT 228  
 QY 50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66  
 Db 229 TTCCCCCGGTGTCTCTACTA-----TGG---TCAGAAAGCCTGTGTGTCTCA 273  
 QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
 Db 274 CCATCTCCAAAGAGGTTACCTGCAGGAAATGTTTAAACGGAGGCTGCCTTCCCTGGCA 333  
 QY 87 ThrArgSerHisLeuGlyArgLysCysSer 97  
 Db 334 ACAAGAGCCACCTGGGAGGAGAAAGTGCAGC 366

RESULT 9  
 US-10-146-502-1307  
 ; Sequence 1307, Application US/10146502  
 ; Publication No. US20030069180A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jiang, Yuqiu  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Stolk, John A.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
 ; FILE REFERENCE: 210121-527C2  
 ; CURRENT APPLICATION NUMBER: US/10/146,502  
 ; CURRENT FILING DATE: 2002-05-14  
 ; NUMBER OF SEQ ID NOS: 2241  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1307  
 ; LENGTH: 572  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 9, 19, 461, 497, 500, 502  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-10-146-502-1307

Alignment Scores:  
 Pred. No.: 6.37 Length: 572  
 Score: 339.00 Matches: 62  
 Percent Similarity: 57.25% Conservative: 13  
 Best Local Similarity: 47.33% Mismatches: 6  
 Query Match: 63.72% Indels: 50  
 DB: 9 Gaps: 28

US-09-854-133-586 (1-97) x US-10-146-502-1307 (1-572)  
 QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
 Db 22 GAGGTGGAG-----AATTGAGAGCAGATGATACACAGAGTGTCTGAGTAGTAATTA 75  
 QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
 Db 76 ---GATGCTGTGAAGGAAAGACACACCTTTGAGTTTTCACCTGTGAACA---CTATAG 129

QY 23 Arg-----LysLys---Glu---Arg---Lys-----LysLys-----ArgGluArg 33  
 Db 130 CGCTGAGAGACAGCTGTAAGAGCAGAGAGATCATCATCAGTACAGTAACCAAGACACACC 189  
 QY 34 LysPheGlnAlaAsnCysGlyLeuAsp---PheIleIlePheTrp-----Ile---Phe 49  
 Db 190 AAA-----GTTGAAAGTTT---GTTTCTTCCCTCTCTCTTTTATTT 228  
 QY 50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66  
 Db 229 TTCCCCCGGTGTCTCTACTA-----TGG---TCAGAAAGCCTGTGTGTCTCA 273  
 QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
 Db 274 CCATCTCCAAAGAGGTTACCTGCAGGAAATGTTTAAACGGAGGCTGCCTTCCCTGGCA 333  
 QY 87 ThrArgSerHisLeuGlyArgLysCysSer 97  
 Db 334 ACAAGAGCCACCTGGGAGGAGAAAGTGCAGC 366

RESULT 10  
 US-09-764-891-6421  
 ; Sequence 6421, Application US/09764891  
 ; Publication No. US20030077808A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PC006  
 ; CURRENT APPLICATION NUMBER: US/09/764,891  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 10231  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 6421  
 ; LENGTH: 1687  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-764-891-6421

Alignment Scores:  
 Pred. No.: 4.25e+03 Length: 1687  
 Score: 298.50 Matches: 58  
 Percent Similarity: 39.78% Conservative: 16  
 Best Local Similarity: 31.18% Mismatches: 10  
 Query Match: 56.11% Indels: 102  
 DB: 9 Gaps: 38

US-09-854-133-586 (1-97) x US-09-764-891-6421 (1-1687)  
 QY 1 GluValGlu-----ValSerArgAspHisAla---SerLeu-----Gly---AspSer 14  
 Db 633 GAGGTAGAGGTTGCGAGTGCAGATCAGCCACCTGCAGCTTCAGCTGGGTGACAGAGC 692  
 QY 15 GluThrLeuSerGlnThrGluLeuArgLysLysGlu-----ArgLys-----Lys 29  
 Db 693 AAGACTTCTCTCAA-----AAAAAAGAGACCTACTCTATAGGAAGCAATTCAG 743  
 QY 30 LysArg---GluArgLysPheGlnAlaAsn-Cys-----Gly----- 40  
 Db 744 AAAAGAGATAAAC---AACTATATG---AATATGTTAGTTTCAGTAGTAATCAGGACAGT 797  
 QY 41 ---Ile-----AspPheIleIle-----PheTrp-----IlePhe 49  
 Db 798 CAAATCAATAATACAGTGAAT---CAAGGTCTCTCTTATTTGTAATAAATAAATTTT 854  
 QY 49 eTrpIleLeuLeu-----PheSer-----His---HisTrp-----IleG 60  
 Db 855 TTTTCTCTTTTGGAGATGCTCTCTCACTTTGTCAACAGGTTGGAGTGCAGTGGTGCA 914  
 QY 60 nGluSer---LeuLeu-----His-----CysProSer--- 68  
 Db 915 A---TCRGCCTCATTCGCAACCTCTGCTCCAGGCTCAATCAGTCTGCCACCTGACC 971

QY 69 -ProLys-----GluVal-----Thr-----Cys----- 74  
Db 972 TCCCAAGTAGCTGTGACTACAGGTCTCCGCCACACCGCTGGCTTAATTTCTTGCATTTT 1031  
QY 75 -----ArgGluMetLeuThrGly-----GlyCysLeuPro---TrpAla---Thr-- 87  
Db 1032 TGATAGAGAT-----GGAGTTTCGCCAGGTTGC---CCAAGCTGGTCTCAAACTCC 1079  
QY 88 -----Arg-----Ser-----HisLeuGly-----ArgArg----- 94  
Db 1080 TGAGCTCAAGCAATCCGCCACCTTGGCTCCCAAGTACTGGGCTTACAGGCGTGAGCC 1139  
QY 95 -----Lys 95  
Db 1140 ACCGAGCCACGACAAA 1156

## RESULT 11

US-09-764-891-6422  
; Sequence 6422, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: ROSEN ET AL.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; PRIOR FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6422  
; LENGTH: 1688  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-6422

## Alignment Scores:

Pred. No.: 4.26e+03 Length: 1688  
Score: 298.50 Matches: 58  
Percent Similarity: 39.78% Conservative: 16  
Best Local Similarity: 31.18% Mismatches: 10  
Query Match: 56.11% Indels: 102  
DB: 9 Gaps: 38

US-09-854-133-586 (1-97) x US-09-764-891-6422 (1-1688)

QY 1 GluValGlu-----ValSerArgAspHisAla---SerLeu-----Gly---AspSer 14  
Db 634 GAGTGGAGTTTCAGTGAGTCGAGATCAGCCACTGCCTTACCTGGGTGACAGAGC 693  
QY 15 GluThrLeuSerGlnThrGluLeuArgLysLysGlu-----ArgLys-----Lys 29  
Db 694 AAGACTTTGTCTCAA-----AAAAAAGACCTACTCTATAGGAAGCAATTCAG 744  
QY 30 LysArg-----GluArgLysPheGlnAlaAsn-Cys-----Gly----- 40  
Db 745 AAAAGAGTAAC---AACTATATG---AATATGTTAGTTTACGTAGTAATTCAGGACGTG 798  
QY 41 -----Ile-----AspPheIleIle-----PheTrp-----IlePh 49  
Db 799 CAAATCAATAAACAGTCAGAT---CAAGTGTCTCTTATTGTTAAATAAATAAATTTT 855  
QY 49 eTrpIleLeuLeu-----PheSer-----His-----HisTrp-----IleGl 60  
Db 856 TTTTCTTTTTCGAGATGCTTCTACTTTGTCACCCAGGTTGGAGTGCAGTGGTGCA 915  
QY 60 nGluSer---LeuLeu-----CysProProSer-- 68  
Db 916 A---TCTGCGCTCATTCACACCTTGCCTCCAGGCTCAATCAGTCTGCCACCTCAGCC 972  
QY 69 -ProLys-----GluVal-----Thr-----Cys----- 74  
Db 973 TCCCAAGTAGCTGTGACTACAGGTCTCCGCCACACCGCTGGCTTAATTTCTTGCATTTT 1032

QY 75 -----ArgGluMetLeuThrGly-----GlyCysLeuPro---TrpAla---Thr-- 87  
Db 1033 TGATAGAGAT-----GGAGTTTCGCCAGGTTGC---CCAAGCTGGTCTCAAACTCC 1080  
QY 88 -----Arg-----Ser-----HisLeuGly-----ArgArg----- 94  
Db 1081 TGAGCTCAAGCAATCCGCCACCTTGGCTCCCAAGTACTGGGCTTACAGGCGTGAGCC 1140  
QY 95 -----Lys 95  
Db 1141 ACCGAGCCACGACAAA 1156

## RESULT 12

US-10-163-866-34  
; Sequence 34, Application US/10163866  
; Publication No. US20030027188A1  
; GENERAL INFORMATION:  
; APPLICANT: EXELIXIS, INC.  
; TITLE OF INVENTION: SLC7s AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE  
; FILE REFERENCE: EX02-080C  
; CURRENT APPLICATION NUMBER: US/10/163,866  
; PRIOR FILING DATE: 2002-06-05  
; PRIOR APPLICATION NUMBER: US 60/296,076  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/328,605  
; PRIOR FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/338,733  
; PRIOR FILING DATE: 2001-10-22  
; PRIOR APPLICATION NUMBER: US 60/357,253  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: US 60/357,600  
; PRIOR FILING DATE: 2002-02-15  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 34  
; LENGTH: 2000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-163-866-34

## Alignment Scores:

Pred. No.: 8.75e+03 Length: 2000  
Score: 296.50 Matches: 57  
Percent Similarity: 58.88% Conservative: 6  
Best Local Similarity: 53.27% Mismatches: 10  
Query Match: 55.73% Indels: 34  
DB: 9 Gaps: 19

US-09-854-133-586 (1-97) x US-10-163-866-34 (1-2000)

QY 2 ValGlu-----ValSerArgAspHisAlaSerLeuGlyAspSerGluThrLeu---SerGln 19  
Db 4 GTGAACACTATA-----GCG---CTG-----AGAGAGACAGCTCTGAAAGCAG 42  
QY 20 ThrGluLeuArgLys-----Lys---GluArgLysLysLysArg---GluArg---LysPhe 35  
Db 43 -----AGGAAGACATCATCATG---AACACCAAGAGACACCAAAAGTTTGAAGTTT 90  
QY 36 GlnAlaAsnCysGlyIleAspPheIle---Ile---PheTrpIlePhe-----TrpIle 51  
Db 91 -----TGT-----TTTCTTCTCCCTCTGTTT---ATTTTCCCGCTGTGTC 129  
QY 52 -LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysProProSerProLysGln 71  
Db 130 CCTACTA-----TGG-----TCAGAAAGCCTGTGTGTCCACCATCTCCAAAGGA 174  
QY 71 uValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAlaThrArgSerHisLe 91  
Db 175 GGTTCCTCAGGGAATGTTAACGGAGGCTGCTTCCCTGGGCAACAAAGAGGACCT 234  
QY 91 uGlyArgArgLysCysSer 97  
Db 235 GGGCAGGAGAAAGTCAGC 253



## RESULT 13

US-09-919-497-9  
; Sequence 9, Application US/09919497  
; Patent No. US2002010662A1  
; GENERAL INFORMATION:  
; APPLICANT: Mutter, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 3995  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-919-497-9

Alignment Scores:  
Pred. No.: 1.19e+05 Length: 3995  
Score: 292.50 Matches: 69  
Percent Similarity: 24.53% Conservative: 9  
Best Local Similarity: 21.70% Mismatches: 11  
Query Match: 54.98% Indels: 229  
DB: 68 Gaps: 68

US-09-854-133-586 (1-97) x US-09-919-497-9 (1-3995)

```
QY 1 GluVal-----Glu-----Val-----SerArg-----Asp----- 7
|||||
Db 2708 GAGATAGATGCGCGGATACCACTTTGTGTCGTCGGGAGAAAATGGAGAGGACATT 2767
QY 8 -----His-----Ala-----Ser-----Leu-----Gly----- 12
|||||
Db 2768 CAGCGGCACAGTTTCATTGAGCGCGCGCCAGTACACAGCCACCTCTATGGGACACGGTC 2827
QY 13 AspSer-----Glu-----Thr-----Leu-----Ser----- 18
|||||
Db 2828 CAGTCCGTGCGAGAGTGGCAGAGCAGGGGAAGCACTGCATCCTCGATGTCTCGGCCAAT 2887
QY 19 -----Gln-----ThrGluLeu-----Arg-----Lys 24
|||||
Db 2888 GCGGTGCGCGGCTGCAGCGCGCCACCTGCACCCCAATCGCCATCTTCATCGGCCCGC 2947
QY 25 Lys-----GluArg-----Lys-----LysArg-----Glu-----ArgLys----- 34
|||||
Db 2948 TCCCTGGAGATGCTAGAGATTAAACAGCGGATCACAGAGGAGCAAGCCGCAAGCC 3007
QY 35 Phe-----Gln-----AlaAsnCys----- 39
|||||
Db 3008 TTCGACAGACCCACCAAGCTGGAGCAGGAGTTTCACAGAGTGCTTCTCAGCCATCGTGAG 3067
QY 40 GlyIleAsp-----Phe-----Ile-----Ile----- 45
|||||
Db 3068 GGT-----GACAGCTTTGAGAGATACACACAGGTGAACGCTGTCTATCGAGGACCTCTCA 3124
QY 46 -----Phe-----TrpIle-----Phe-----TrpIle-----Leu 52
|||||
Db 3125 GCGCCCTACATCTGGTCTCCAGCCCGGAGAGACTCTGATCTCTGCGCTGGCTGGCTG 3184
QY 53 -----LeuPheSerHisHis-----TrpIleGlnGluSer-----LeuLeu 64
|||||
Db 3185 GACTCGCCCTGCGCTC-----CATCACCTGGGCGCCCTGG-----TCTGACATGAAT 3229
QY 65 Cys-----ProPro-----Ser-----Pro----- 69
|||||
Db 3230 TGCCCAAGCCCTTGGCTCCCGCGGCTCCCTCCACCCCTTCTTATTTATTTCTTTCT 3289
QY 70 Lys-----Glu-----Val-----Thr-----Cys-----Arg-----Glu 76
|||||
```

```
Db 3290 AACTGATCCAGCCTTTGGAGGGGGGACACTCTCTGCATGTATCCCGCACCCAGAA 3349
QY 77 MetLeu-----Thr-----GlyGly-----Cys----- 82
|||||
Db 3350 ---CTGGGCTCTGAACGCGCAGGAACCTGGGGTCTGGGGGGAGCTGGGCTCCTTGTTC 3406
QY 83 -----Leu-----ProTrpAla----- 86
|||||
Db 3407 GAGCCCTTGTCTCTTAGATCCCC---GCCCCACCTGCCCCCAATGCACACACAGACC 3463
QY 87 -----Thr-----ArgSer----- 89
|||||
Db 3464 ACCGGGGGCCACCTGCGCTCCGCCATCTCTCCACACACATTCAGAGAGTCCAGGGCCCC 3523
QY 90 -----His-----Leu-----Gly-----Arg----- 94
|||||
Db 3524 CTCGAGGAGACCCCGCTGTCAGGGATGCAGGGCCACAGGCCCTCCGCTCTCTCTTAAGCAG 3583
QY 95 -----Lys-----Cys-----Ser 97
|||||
Db 3584 GGTCTGGGGTCAACCCCTGCTCATCGTAATTCGCCATGTTACCTTGATTCTCA 3637
```

## RESULT 14

US-09-254-783A-2  
; Sequence 2, Application US/09254783A  
; Patent No. US20020035734A1  
; GENERAL INFORMATION:  
; APPLICANT: Communil, Didier  
; APPLICANT: Boeynaems, Jeanmarie  
; TITLE OF INVENTION: G Protein Coupled Receptor Showing Selective Affinity for  
; FILE REFERENCE: 9409/2062  
; CURRENT APPLICATION NUMBER: US/09/254,783A  
; CURRENT FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: PCT/BE98/00108  
; PRIOR FILING DATE: 1998-07-09  
; PRIOR APPLICATION NUMBER: EP97870101.9  
; PRIOR FILING DATE: 1997-07-09  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 2427  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-254-783A-2

Alignment Scores:  
Pred. No.: 3.13e+04 Length: 2427  
Score: 287.50 Matches: 63  
Percent Similarity: 30.36% Conservative: 12  
Best Local Similarity: 25.51% Mismatches: 17  
Query Match: 54.04% Indels: 155  
DB: 10 Gaps: 56

US-09-854-133-586 (1-97) x US-09-254-783A-2 (1-2427)

```
QY 1 GluValGlu-----Val-----SerArgAspHisAlaSer-----LeuGly----- 12
|||||
Db 1120 GAA---GAGGCGCTCTGGGATCCCTTCAAGACG---CCGAGCCTGGAGTTGGTGAGGAC 1173
QY 13 ---Asp---SerGlu-----ThrLeuSer-----Gln----- 19
|||||
Db 1174 GATGATGAACAGGAAGATGATGATCGATATTTCTGCCAGCGGTGGCGAGGCGCC 1233
QY 20 ThrGlu---Leu-----Arg---Lys-----LysGlu-----ArgLys 28
|||||
Db 1234 AGTGAGGACCTGTTCCCGGAGGCCAAGCAAGCGGCTGCCAAGTCTCCAGGGCGGAAG 1293
QY 29 LysLysArg---Glu-----Arg-----Lys-----Phe-----Gln 36
|||||
Db 1294 CGGAAGCGGTGGGAATGATGATGAGGTGCCAGTCTCCCTGCCAAGTCTTGGCAGCT 1353
QY 37 Ala---Asn-----CysGlyIle-----AspPheIlePheTrp---Ile----- 48
|||||
```

Db	1354	GCCGACGACAAATCAGTGGTTCCAGGGGAGCTTC	---CTG---TGGCCCACTACTGGTG	1407
QY	49	-----PheTrp-----	Leu-----PheSer-----	55
Db	1408	GTTGAGTTCTGGTGGCGGTGCCAGCAATAGCGCCTGTACCGCTTCAGCATCCGG		1467
QY	56	-----HisHis---Trp-----	-----IleGln---Glu---Ser---Leu63	
Db	1468	AAGCAGCGCCCATGCGACCCCGCGTGGTCTCTCTGTGCCACTGGCAGCTCAGCGACCTG		1527
QY	64	LeuCys-----	-----ProPro-----Ser-----ProLys-----	70
Db	1528	CTCTGGCGTCTGACGCTCGCCCGGTGGCGGCTACCTTACTTCCGCCCAAGCACTGGCGC		1587
QY	71	-----GluValThrCysArg---Glu---MetLeu---Thr-----	glyGly81	
Db	1588	TATGGGAGGCGGGTGGCGCTGGAGGCTTCCTCTTCACCTGCACCTGCTGGGCAGC		1647
QY	82	-----Cys-----		82
Db	1648	GTCATCTTCATCACTGCATCAGCCTCAACCGCTACCTGGGCATGTCGACCCCTTCTTC	1707	
QY	83	-----Leu-----Pro-----	-----Trp-----	85
Db	1708	GCCGGAAGCACCCTGCGACCCCAAGCAGCGCTGGGCCGTGAGCGCTGCGCGGTGGTCTG	1767	
QY	86	-----Ala-----Thr---Arg---SerHisLeuGlyArgArg-----	94	
Db	1768	GCGCGCTGTGCGCATGCCCACACTCAGCTTCTCCACCTG---AAGAGGCCGCCGAG	1824	
QY	95	Lys-----CysSer97		
Db	1825	CAGGGGCGGGCACTGCAGC	1845	

```

RESULT 15
US-10-152-058-2
; Sequence 2, Application US/10152058
; Patent No. US20020142988A1
; GENERAL INFORMATION:
; APPLICANT: Communi, Didier
; APPLICANT: Boesynaems, Jeanmarie
; TITLE OF INVENTION: G protein Coupled Receptor Showing Selective Affinity for ATP
; FILE REFERENCE: 9409/2062
; CURRENT APPLICATION NUMBER: US/10/152,058
; CURRENT FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: PCT/BE98/00108
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: EP97870101.9
; PRIOR FILING DATE: 1997-07-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-152-058-2

```

Alignment Scores:		
pred. No.:	3.13e+04	Length:
Score:	287.50	Matches:
Percent Similarity:	30.36%	Conservative:
Best Local Similarity:	25.51%	Mismatches:
Query Match:	54.04%	Indels:
DB:	12	Gaps:
		2427
		63
		17
		155
		56

US-09-854-133-586 (1-97) x US-10-152-058-2 (1-2427)

QY	1	GluValGlu-----Val--SerArgAspHisAlaSer-----LeuGly-----	12
		::	
Db	1120	GAA--GAGGCGCTCTGGGATCCCTTCAGGACG--CGAGCCTGGAGTTGGGTAGGAC	1173
QY	13	Asp--SerGlu-----ThrLeuSer-----Gln-----	19
		::	

Db	1174	GATGATGAACAGGAAGATGATGACATCGAGTATTCTCCAGGCGGTGGCGAGGCGCCC	1233
Qy	20	ThrGlu---Leu-----Arg---Lys-----LysGlu-----ArgLys	28
Db	1234	AGTGAGGACCTGTTCCCGGAGGCCAAGCAAAAGCGCTTGCCAAGTCTCCAGGCGCGAAG	1293
Qy	29	LysLysArg---Glu-----Arg-----Lys-----Phe-----Gln	36
Db	1294	CGGAAGCGGTGGGAATATGGATCGAGGTGCCAAGTCTGCCCTGCCAACTCTTTGGCAGCT	1353
Qy	37	Ala---Asn-----CysGlyIle-----AspHeIleIlePheTrp---Ile-----	48
Db	1354	GCCGACGACAACTAGTGGGTTCACGGGGACTTC---CTG---TGGCCCATACTGGTG	1407
Qy	49	-----PheTrp-----Ile-----Leu-----Leu-----PheSer-----	55
Db	1408	GTTGAGTCTCTGTGTGGCCGTGGCCGCAATGGCCTGGCCCTGTACCGCTTCAGCATCCGG	1467
Qy	56	---HisHis---Trp-----IleGln---Glu---Ser---Leu	63
Db	1468	AGCAGCGCCATGGGACCCCCCGGTGTCTCTGTCTCCAGCTGGCAGTCAGCAGCACCTG	1527
Qy	64	LeuCys-----ProPro-----Ser-----ProLys-----	70
Db	1528	CTCTGCGCTCTCAGCGTGCCTGCCCGCGTGGCCGCTACCTCTATCCCCCAAGCACTGGCGC	1587
Qy	71	-----GluValThrCysArg---Glu---MetLeu---Thr-----GlyGly	81
Db	1588	TATGGGGAGCGCGTGGCGCTGGAGCGCTTCCTCTTCAGCTGCACACCTGCTGGGCAGC	1647
Qy	82	-----Cys-----Leu-----Leu-----Leu-----	82
Db	1648	GTCATCTTCATCACCCTGCATCAGCCTCAACCGCTACCTGGGCGATCGTGCAACCCCTTCTTC	1707
Qy	83	-----Leu---Pro-----Leu-----Leu-----Leu-----	85
Db	1708	GCCGAAAGCCACCTGCAGACCCCAAGCAGCGCTGGGCCGTGAGCGCTGCGGGCTGGGTCTG	1767
Qy	86	-----Ala-----Thr---Arg---SerHisLeuGlyArgArg-----	94
Db	1768	GCGCGCCCTGCTGGCCATGCCACACTCAGCTTCTCCACACTG---AAGAGGCGCGCCGAG	1824
Qy	95	Lys-----CysSer	97
Db	1825	CAGGGGCGGGCACTGCAGC	1845

Search completed: May 11, 2003, 20:09:26  
Job time : 94 secs

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

SUMMARIES

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 17:46:32 ; Search time 1056 Seconds  
(without alignments)  
1487.654 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVRDASLGDSLTSLQT.....LFGGCLPWATRSHLGRKKS 97

Scoring table:  
BLOSUM62 0.0 , Xgapext 0.5  
Ygapop 0.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0  
Maximum DB seq length: 5000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n\_model -DEV=xlh  
-O=/Cgn2\_1/USP001/US09854133/runat\_05052003\_174426\_1224/app\_query.fasta\_1.263  
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=5000  
-USER=US09854133 @CGN\_1\_1\_1456 -Bunat\_05052003\_174426\_1224 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAL -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=0 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=0 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_man.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_fod.\*

Result No.	Score	Match	Length	ID	Description
1	339.5	63.8	936	12	BG284503
2	334.5	62.9	1072	12	BG388107
3	307.5	57.8	1143	12	BG260656
4	305	57.3	557	17	AQ002318
5	299.5	56.3	987	12	BG248967
6	294.5	55.4	1300	12	BG286262
7	293.5	55.2	1049	13	BM478521
8	293	55.1	882	12	BF699490
9	292.5	55.0	910	12	BG326527
10	290	54.5	937	12	BG391278
11	288.5	54.2	1062	10	BE035217
12	287	53.9	862	14	BQ688218
13	287	53.9	876	14	BQ687348
14	286.5	53.9	748	17	AZ740600
15	286.5	53.9	1021	12	BG261332
16	286.5	53.9	1051	12	BG026259
17	286	53.8	913	12	BF696158
18	285.5	53.7	960	17	CNS03MEM
19	285	53.6	597	17	BH787408
20	285	53.6	836	10	BE538494
21	285	53.6	956	12	BG169397
22	285	53.6	998	12	BE793114
23	285	53.6	1119	17	CNS06QSG
24	285	53.6	1210	13	BM561270
25	285	53.6	1242	17	B12259
26	284.5	53.5	1006	14	BQ728437
27	284	53.4	1310	11	AF116717
28	284	53.4	1780	12	BG115708
29	283.5	53.3	1092	17	CNS038E2
30	283	53.2	980	10	BF570665
31	283	53.2	960	17	CNS041ML
32	283	53.2	989	17	CNS03D07
33	283	53.2	1136	14	BM928444
34	282.5	53.1	936	13	B1199098
35	282.5	53.1	1038	12	BE913378
36	282.5	53.1	1271	13	BI687475
37	282	53.0	839	10	BE641384
38	282	53.0	920	13	BM465182
39	282	53.0	943	12	BG536746
40	282	53.0	1769	12	BG170413
41	282	53.0	2917	11	AK019171
42	281.5	52.9	835	12	BG253649
43	281.5	52.9	853	17	AZ740605
44	281.5	52.9	886	12	BF215972
45	281.5	52.9	1023	12	BE797979

ALIGNMENTS

RESULT 1  
BG284503  
LOCUS 602408645f1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4537810 5',  
DEFINITION mRNA sequence.  
ACCESSION BG284503  
VERSION BG284503.1 GI:13035516  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 936)  
AUTHORS NIH-MGC <http://mgc.mci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
COMMENT

Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DP  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 Clone sequencing by: Incyte Genomics, Inc.  
 found through the I.M.A.G.E. Consortium information can be  
 http://image.llnl.gov  
 Plate: LLAM10463 row: i column: 11  
 High quality sequence stop: 795.  
 Location/Qualifiers  
 1..936  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4537810"  
 /clone\_lib="NIH\_MGC\_91"  
 /tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: prostate; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dr primed.  
 Average insert size 1.4 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 248 a 202 c 251 g 235 t

BASE COUNT  
ORIGIN

## Alignment Scores:

Pred. No.: 356 Length: 936  
 Score: 339.50 Matches: 62  
 Percent Similarity: 57.25% Conservative: 13  
 Best Local Similarity: 47.33% Mismatches: 7  
 Query Match: 63.82% Indels: 49  
 DB: 12 Gaps: 28

US-09-854-133-586 (1-97) x BG284503 (1-936)

QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
 Db 41 GAG---GAGGTGGAGATTGAGACGACGATGCATACACAGGTGTTCTGAGTAGTAATTA 97  
 QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
 Db 98 ---GATCGCTGAGAGGAAAGACACCTTTGAGTTTCCCTGTGAACA---CTATAG 151  
 QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
 Db 152 CGCTGAGAGACAGCTGAAAGCAGAGAGATCGATCAGTAACACCAAGACACACC 211  
 QY 34 LysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp-----Ile---Phe 49  
 Db 212 AAA-----GTTGAAAGTTT---GTTTCTTCTCCCTGTGTTTATTT 250  
 QY 50 Trp-----Ile---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66  
 Db 251 TTCCTCCCTGTGTCCTACTA-----TGG---TCAGAAAGCCTGTGTGTCTCCA 295  
 QY 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
 Db 296 CCATCTCAAGAGGTTTACCTGCGAGGAAATGTTAACGGAGGCTGCCCTCCCTGGGCA 355  
 QY 87 ThrArgSerHisLeuGlyArgArgLysCysSer 97  
 Db 356 ACAAGGACCCCTGGCAGGAGAAAGTGCAGC 388

## RESULT 2

BG388107  
 LOCUS  
 DEFINITION 602413070F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4521736 5',  
 mRNA sequence.  
 BG388107  
 ACCESSION  
 VERSION BG388107.1 GI:13281553

KEYWORDS  
SOURCE  
ORGANISM

EST.  
human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1072)

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone sequencing by: Incyte Genomics, Inc.

found through the I.M.A.G.E. Consortium information can be

http://image.llnl.gov

Plate: LLAM10421 row: k column: 17

High quality sequence stop: 625.

Location/Qualifiers

1..1072

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4521736"

/clone\_lib="NIH\_MGC\_92"

/tissue\_type="embryonal carcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally; oligo-dr primed.

Average insert size 2.5 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC Library."

331 a 229 c 270 g 242 t

BASE COUNT  
ORIGIN

## Alignment Scores:

Pred. No.: 990 Length: 1072  
 Score: 334.50 Matches: 62  
 Percent Similarity: 57.36% Conservative: 12  
 Best Local Similarity: 48.06% Mismatches: 10  
 Query Match: 62.88% Indels: 45  
 DB: 12 Gaps: 28

US-09-854-133-586 (1-97) x BG388107 (1-1072)

QY 1 GluValGluValSerArg---AspHis---Ala-----Ser-----Leu 11  
 Db 59 GAG---GAGGTGGAGATTGAGACGACGATGCATACACAGGTGTTCTGAGTAGTAATTA 115  
 QY 12 GlyAsp---SerGlu-----Thr---LeuSer---Gln-----ThrGluLeu--- 22  
 Db 116 ---GATCGCTGAGAGGAAAGACACCTTTGAGTTTCCCTGTGAACA---CTATAG 169  
 QY 23 Arg-----LysLys---Glu---Arg---Lys---LysLys-----ArgGluArg 33  
 Db 170 CGCTGAGAGACAGCTGAAAGCAGAGAGATCGATCAGTAACACCAAGACACACC 229  
 QY 34 LysPheGlnAlaAsnCysGlyIleAspPhe-----IleIlePheTrpIlePhe--- 49  
 Db 230 AAAAGTTGAA---AGTTTA---GTT---TCTTCTCCCTGTGTTTATTT---TTCCTC 274  
 QY 50 TrpIle---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysProSer 68  
 Db 275 CGTGTGTCCTACTA-----TGG---TCAGAAAGCCTGTGTGTGTCTCCA 319  
 QY 69 ProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAlaThrArg 88  
 Db 320 CCAAGAGAGGTTTACCTGCGAGGAAATGTTAACGGAGGCTGCCCTCCCTGGGCAACAAG 379  
 QY 89 SerHisLeuGlyArgArgLysCysSer 97  
 Db 380 AGCCACCTGGCAGGAGAAAGTGCAGC 406

us-09-854-133-586.rst

Tue May 13 12:12:51 2003

RESULT 3  
 BG260656  
 LOCUS  
 DEFINITION  
 60237222F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4480277 5',  
 mRNA sequence.  
 1143 bp linear EST 13-FEB-2001  
 BG260656  
 ACCESSION  
 BG260656.1 GI:12770472  
 KEYWORDS  
 EST.  
 SOURCE  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1143)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cygaps@remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM10313 row: 1 column: 06  
 High quality sequence stop: 709.  
 Location/Qualifiers  
 1. 1143  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4480277"  
 /clone\_lib="NIH\_MGC\_93"  
 /tissue\_type="transitional cell papilloma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: bladder; Vector: PCW-SPORT6; Site: 1: NotI;  
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.7 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 371 a 287 c 231 g 254 t  
 BASE COUNT  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.18e+04 Length: 1143  
 Score: 307.50 Matches: 68  
 Percent Similarity: 28.08% Conservative: 14  
 Best Local Similarity: 23.29% Mismatches: 7  
 Query Match: 57.80% Indels: 203  
 12 Gaps:  
 DB:  
 US-09-854-133-586 (1-97) x BG260656 (1-1143)  
 QY 1 GluValGlu-----ValSerArgAspHisAla-----SerLeu-----Gly----- 12  
 Db 36 GAGCGGAGCTTGCAGTGAGCGGAGATCAGCGGAGTGCAGTGCCTCAATGCGTGGCGGACAG 95  
 QY 13 ---Asp-----Ser-----Glu-----Thr-----Leu-----SerGlnThr 20  
 Db 96 CGAGACTCCGGGAGCGGAGCTTGCAGTGCAGCGGAGATCAGCGGAGTGCAGTCA 152  
 QY 21 -----Glu-----LeuArg-----LysLysGluArgLysLysLys-----Arg 31  
 Db 153 CTGGCGGACAGCGGAGCTCCGCTCAAAAAA-----AAAAAAGAAATTTCTTACTAGA 209  
 QY 32 GluArgLys-----PheGln-----AlaAsnCys----- 39  
 Db 210 GAG---AAACCCCTTACAAATTTTAAAGAAATTTGGCAAAATTTCTTAAACCGGCTTCAAC 266  
 QY 40 -----Gly-----Ile-----Asp-----Phe--- 43  
 Db 267 CTTTACTTAACATTAGGTAATTTCTATCTAGTGGAGTTAAACTTCTTCAATGTTTGC 326

QY 44 ---IleIle-----Phe---Trp-----Ile--- 48  
 Db 327 CGAATTGTGGCAAAATCCTTTTCTCTGGTCTCAGCCCTAACCGACCTTAAGAAATTTCA 386  
 QY 49 PheTrp-----Ile-----LeuLeu-----Phe 54  
 Db 387 TACTGGACAGCAACCCCTTACAACAGGAGACACATTTTGGCAAAAGCCCTTTTAAACACGATTT 446  
 QY 55 SerHis-----His----- 57  
 Db 447 ---CACACCTTTTACAGATAAGATAACTCATATTTGGAGAGAATCTTACAAGTGTGAA 503  
 QY 58 -----Trp-----IleGln-----Glu-----Ser 62  
 Db 504 CTAATGTGTACACAGCTCTATAAAAGTCTCTGAATTTCTTAAACAGACATTAATATTATCA 563  
 QY 63 -----Leu-----LeuCysProProSerPro----- 69  
 Db 564 TACTGGAGAGAAACTCTAAACACCTGAAAGATGTTG---CCACCG---CCTTTTGACAAG 617  
 QY 70 -----Lys-----Glu-----ValThr 73  
 Db 618 TCAAACTTTTCAACATTCACGAAATCATGCTGTGAGAAATCTTACAGACATTAATATTATCA 677  
 QY 74 ---Cys-----ArgGlu-----Met 77  
 Db 678 ATTTGTGCACCAAGCCCTCTAAATGATGTCCTTGTGCTAGCTAGCAAAATTCATA 737  
 QY 78 LeuThrGly---Gly-----CysLeuPro-----Trp 85  
 Db 738 CTG---GGGAAATATTCACCGGTGTGTCCACCCACTGTGTGCCACCCCTTTTGACCTATTGG 794  
 QY 86 Ala---Thr-----Arg-----Ser----- 89  
 Db 795 TCCCGCACCTTAAAGAAATGGAACGCTTATGCTGCAGACACATGTGCACACTATGTA 854  
 QY 90 His-----Leu-----GlyArgArgLysCys---Ser 97  
 Db 855 CACACCTTATATGCGGG---TCCAAGTGCCTCACT 887  
 RESULT 4  
 AQ002318/c  
 LOCUS  
 DEFINITION  
 sequence.  
 ACCESSION  
 AQ002318  
 VERSION  
 AQ002318.1 GI:3029522  
 KEYWORDS  
 GSS.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 557)  
 Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,  
 Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,  
 Simon, M., and Venter, J.C.  
 Use of a random BAC End Sequence Database for Sequence-Ready Map  
 Building (1998)  
 Unpublished (1998)  
 Other\_GSSs: CIT-HSP-2283E8.TR  
 CONTACT: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: M13-21;  
 Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. .557  
/organism="Homo sapiens"  
/db\_xref="GDB:7147907"  
/db\_xref="taxon:9606"  
/clone="228388"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2: HindIII"  
BASE COUNT 142 a 152 c 133 g 130 t  
ORIGIN

## Alignment Scores:

Pred. No.: 551 Length: 557  
Score: 305.00 Matches: 557  
Percent Similarity: 57.52% Conservative: 10  
Best Local Similarity: 48.67% Mismatches: 10  
Query Match: 57.33% Indels: 38  
DB: 17 Gaps: 22

US-09-854-133-586 (1-97) x A0002318 (1-557)

QY 7 Asph1aLaserLeuGlyAspSerGluThr---LeuSer---Gln-----ThrGluLeu 22  
Db 557 GATCAGCTGNA---GGATATAGC---ACACCTTTGAGTTTTCACCTGTGAACA---CTA 507  
QY 23 ---Arg---Lys---Glu---Arg---Lys---LysLys-----ArgGlu 32  
Db 506 TAGCGCTGAGATACAGTCTGAACGAGAGTAGACATCGATCAGTAACACCAAGAGAC 447  
QY 33 ArgLysPheGlnAlaAsnCysGlyIleAsp---PheIleIlePheTrp---Ile 48  
Db 446 ACCAAA-----GTTGAAAGTTT---GTTTCTTCTCCTCTGATTTA 408  
QY 49 PheTrpIleLeuLeuPheSerHisHis-----TrpIleGlnGluSerLeuLeu 64  
Db 407 TTT-----TTC---CACCGTGTGCTCCCTACTATGG---TCAGAAAGCCTGTG 366  
QY 65 CysProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuPro 84  
Db 365 TGTCCACCATCTCCAAAGTAGTTACCTGCAGGGAATGTTAACGGAGGCTGCCCTCCC 306  
QY 85 TrpAlaThrArgSerHisLeuGlyArgLysCysSer 97  
Db 305 TGGCACACAGGAGGCCACCTGGGCAGCAGAAAGTGCAGC 267

RESULT 5  
BG248967

LOCUS  
DEFINITION BG248967 987 bp mRNA linear EST 13-FEB-2001  
602361925F1 NIH\_MGC\_89 Homo sapiens CDNA clone IMAGE:4470054 5',  
mRNA sequence.  
ACCESSION BG248967  
VERSION BG248967  
KEYWORDS EST.  
SOURCE human.  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 987)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: L1AM10287 row: b column: 07  
High quality sequence stop: 493.  
FEATURES  
source  
1. .987  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4470054"  
/clone\_lib="NIH\_MGC\_89"  
/tissue\_type="hypernephroma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: PCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 274 a 299 c 229 g 185 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.16e+04 Length: 987  
Score: 299.50 Matches: 64  
Percent Similarity: 28.99% Conservative: 16  
Best Local Similarity: 23.19% Mismatches: 9  
Query Match: 56.30% Indels: 187  
DB: 12 Gaps: 56

US-09-854-133-586 (1-97) x BG248967 (1-987)

QY 1 GluValGlu-----ValSerArgAspHisAla---SerLeu-----Gly---AspSer 14  
Db 64 GAGTGGAGGTTCAGTGCAGCGAGATCATGCCACTGCATCCAGCCTGGGCAACAGAGT 123  
QY 15 GluThrLeuSerGlnThrGluLeuArgLysLysGluArgLys-----Lys--- 29  
Db 124 GAGACTTTGTCTCAAGAAAA---AAAAAATAAATAATCTGGGGGATGAAATAA 180  
QY 30 ---Lys---Arg---Glu-----ArgLys-----Phe 35  
Db 181 CCATAATGAAATAGTAGTAAATACTCAGCAAGCAGGAGCGCTCTTCTCACCCTTT 240  
QY 36 -----Gln-----Ala----- 37  
Db 241 GGTCCCTCGGGGATAAATCATCAGTCACTATTACAACAACACTGAAAGCCTCTGCTTAA 300  
QY 38 Asn-----Cys-----GlyIleAspPheIle----- 44  
Db 301 CACCAATGATGCACAGCAGTACTTACTTATGTATTTCCGAGT---GACCTATATAGATG 357  
QY 45 -----IlePheTrp-----Ile---Phe-----Trp-----Ile 51  
Db 358 GAGAAGAGGTGCATTGGGCAGAACTGATGTATAGGAGAGCCCATTCAGCTTCTC 417  
QY 52 -----Leu-----LeuPheSerHis-----His 57  
Db 418 CCTCGAGGAGCAAAATGCTCTCGGTCAATTATAGCTCTTTAC---CACAGACTGGGCCAT 474  
QY 58 -----Trp----- 58  
Db 475 TCTAGTGAACCGGTGGGTACCCCTTGGGGTGGCCCAAGAAACCAATCTTATCTGTTT 534  
QY 59 -----Gln-----Ser-----LeuLeuCys----- 65  
Db 535 CCACCTACCTCAGGAGCGAAGCAGTACGCGGGTCAATCCAAATC---TGATTTGCAAGC 591  
QY 66 -----Pro---ProSer---Pro---LysGluValThrCysArgGluMet 77  
Db 592 CTATCTCATCTCCTCCCAACCTTCTCCTCTGGGAAACCC---AACTGC-----CTA 642  
QY 78 -----LeuThr---Gly-----Cys-----Leu-----Pro--- 84  
Db 643 TCTCATCTGACCAAGAGCTTAGCAGGAGTAAACGCTGACAGACTCGCAGCGGCCCTAT 702  
QY 85 -----Trp-----Ala-----Thr----- 87

```

Db 703 CGCGCTGGGGGAGGCCCTTTGCTAAACCCAGGAAGCGCGCGGACTAAACAAGG 762
QY 88 Arg-----Ser---His-----Leu---Gly 92
Db 763 AGGCCCCAAGCGGGCGCGGATCCAGCACTCCAGGCCCGCCCGGGAACGAGCGT 822
QY 93 -----ArgArg-----Lys-----Cys 96
Db 823 TGCACACACCTGGCTAAAGGAGCGCTATCCAGCGCTCTCTCCCGCTGC 870

RESULT 6
BG286262 1300 bp mRNA linear EST 21-FEB-2001
LOCUS 602383093F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4500599 5',
DEFINITION mRNA sequence.
ACCESSION BG286262
VERSION BG286262.1 GI:13038979
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1300)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10366 row: j column: 24
High quality sequence stop: 366.
FEATURES
source
1..1300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4500599"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: bladder; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 287 a 390 c 313 g 309 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 6.08e+04 Length: 1300
Score: 294.50 Matches: 66
Percent Similarity: 29.8% Conservative: 12
Best Local Similarity: 29.2% Mismatches: 12
Query Match: 55.36% Indels: 171
DB: 12 Gaps: 50

US-09-854-133-586 (1-97) x BG286262 (1-1300)
QY 1 Glu-----ValGluValSer---ArgAspHisAla---SerLeu-----GlyAsp--- 13
Db 180 GAGCGGGGGGTTCCAGTGAAGTAAAT---CATGCCACTGCATCCAACTGGGACAG 236
QY 14 SerGluThrLeuSerGlnThrGluLeuArgLysGluArgLysLys---Lys----- 30
Db 237 AGTGAACCTCTGTCTCAA---AAA---AAAAAAAAAAAAAAAAAAATCCAAATTCGG 290
QY 31 ---ArgGluArg-----Lys---PheGlnAlaAsnCys---GlyIleAspPhe 43

```

```

Db 291 GTGCGGAAAGATGGGAAAAAAACCTTT---TTGGGTGTGATGGGACCCATT 347
QY 44 Ile-----Phe-----TrpIlePheTrp-----IleLeu 52
Db 348 GTGTTGAAATTTTTTCTCCGTGCAGGTGCTGG-----TGTTCCCACTTCTT 401
QY 53 -----Leu-----PheSerHis-----HisTrp----- 58
Db 402 ATTTACCCCTCCNCTATTTCTCTCATTTCTCATCTCATATTTCTTTTATAA 461
QY 59 -----Ile-----Gln----- 60
Db 462 ACCCTATACCTATGTGTGTATAGGCGCACAGGTTCTCTTAAAGAGACCTTGG 521
QY 61 -----Glu-----Ser-----LeuCys----- 65
Db 522 AGAATCCCGCGGGGAGAGAACTAAGAGCCCTCTGTGTCTCTCGTGTGGAGCC 581
QY 66 Pro-----Val---Thr-----Pro---Ser---ProLysGlu--- 71
Db 582 CTTTGTGTTGTTCTCTTAGCGGGGGGACACCCACGCTGGGAGACCC---GAT 638
QY 72 -----Val---Thr-----Cys----- 74
Db 639 ACACCTTTTGTGTTCACTCCCTCTGTGTGCTCGGCTCTGTCTCTCACAGG 698
QY 75 ---ArgGluMet---Leu-----Thr----- 79
Db 699 CTCGCGAGGTGGCCCTTTCTCGCCACCGCTTTTAGCCACACGCGGGCGGCA 758
QY 80 -----Gly-----Gly-----CysLeu 83
Db 759 CCGCAGCGGGGGGCAACTCGTGTCTCTGTGCGGGGGCGCGCGCGGTGTCTC 818
QY 84 ---Pro---Trp-----Ala---Thr-----ArgSerHisLeu----- 91
Db 819 CCCCCGAGATGCTTGGGCCATCTCGGCATTCACACCTCTGCGCTCTATTGTAG 878
QY 92 -----GlyArg-----LysCys 96
Db 879 CCCCAGGAGATCTTTCTCCGAGACAAATTTCTTTGGTGGACGCCCCCACT 938
QY 97 Ser 97
Db 939 GCC 941

RESULT 7
BM478521 1049 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT.6457605 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5575203
DEFINITION 5' mRNA sequence.
ACCESSION BM478521
VERSION BM478521.1 GI:18527563
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1049)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12325 row: b column: 04

```

FEATURES High quality sequence stop: 705.  
source Location/Qualifiers  
1. 1049

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5575203"  
/clone\_lib="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: PCW-SPO16; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 286 a 253 c 240 g 262 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2.48e+04 Length: 1049  
Score: 293.50 Matches: 67  
Percent Similarity: 27.84% Conservative: 14  
Best Local Similarity: 23.07% Mismatches: 11  
Query Match: 55.17% Indels: 199  
DB: 13 Gaps: 61

US-09-854-133-586 (1-97) x BM478521 (1-1049)

```

Qy 1 GluValGlu-----ValSerArg-----Asp-----7
|||||
Db 186 GAGGTGGAGGGAAGATGCTAGTAGGACAGAGGTAACTATGATGACTGCTCATTTGT 245
Qy 8 ---HisAlaSerLeu-----GlyAsp---SerGlu-----Thr---16
|||||
Db 246 GGAATGCTCCGCTGAAGGCCCGGAGGACAGGACAGGACAGTATCACTATCATGTGT 305
Qy 17 -----LeuSerGln-----Thr---Glu---Leu 22
|||||
Db 306 GTCGATATGTCATGAACCATCACTG---CAGGAACAAGTTTCACAAAAGAACCCATAC 362
Qy 23 ArgLys-----LysGlu---ArgLys-----Lys-----Arg---31
|||||
Db 363 AAGAAGTACATCAAGATATACATCAATCAAGGGAACCTTGAAGACAGAGACCA 422
Qy 32 GluArg---Lys---Phe-----Gln-----Ala---37
|||||
Db 423 GAAAGAGTAAACCTTTATGACAGGGGCTGCAGAACAAATCAAGCACATCCTTGCTAAT 482
Qy 38 -----AsnCys-----Gly---Ile-----41
|||||
Db 483 TTCAAAACTACAGTCTTTATTTGTTGAACACATGATCCAGATGGCATGTTGCTCTA 542
Qy 42 ---AspPhe-----Ile-----Ile---Phe-----46
|||||
Db 543 TTGGACTACCGTGAAGATGGTGTGACCCCATATATATTTCTTAAAGGATGTTAGAA 602
Qy 47 -----Trp---IlePheTrpIle-----Leu-----53
|||||
Db 603 ATGGAATAATGTTAAACAATGGCAATTTATTTGGATCTATCATCCTGTATCAATACTG 662
Qy 54 PheSer-----His---His-----Trp---IleGlnGluSer 62
|||||
Db 663 GCTTCTGTTGTCATCCACACACACAGGACCTTAAGACAAATGGGACTGATGTCATCT 722
Qy 63 -----Leu-----Cys-----65
|||||
Db 723 TGAGCTCTTCATTTATTTGACTGTGATNTATTTGGAGTGAGGACTGTTTAAAGAA 782
Qy 66 -----Pro-----Ser-----68
|||||
Db 783 AACATGCATGTAGTTGCCCTAAATAAATGATTTAACTCCNNNNNNNAAAAACCCC 842
Qy 69 -----ProLys-----GluValThrCys---Arg---GluMet-----77
|||||

```

```

Db 843 CCCACATNCCCAACATTTCTCA---ACGGCGGGCCGCTCTAAATATCCCTCGGAGG 899
Qy 78 -----LeuThr-----Gly---GlyCys-----Leu---Pro-----Trp 85
|||||
Db 900 GGTCCTCCAGCTTACCCCGGTAAACCGGCTTTCTTGTACAAAGGTGGGCCCTATTATTGG 959
Qy 86 -----Ala-----Thr-----Arg---88
|||||
Db 960 GGCGGGTTTATACCGCTTAGTCACGTGCGCGCTCTTTTATACAAACCGGCTCGTGT 1019
Qy 89 SerHisLeuGlyArgLys---Cys---Ser 97
|||||
Db 1020 TCT-----GGG---GGAACAAACATGCCTAACCC 1043
RESULT 8
BF699490 882 bp mRNA linear EST 22-DEC-2000
LOCUS 602126079F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4282962 5',
DEFINITION mRNA sequence.
ACCESSION BF699490
VERSION BF699490.1 GI:11984898
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 882)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM113 row: n column: 19
High quality sequence stop: 672.
FEATURES Location/Qualifiers
1. 882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4282962"
/clone_lib="NIH_MGC_56"
/tissue_type="primitive neuroectoderm"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccctcgcc); Site_2: SfiI (ggccattggcc);
Double-stranded cDNA was prepared from cell line RNA. 5',
and 3' adaptors were used in cloning as follows: 5',
adaptor sequence: 5'-ATTCTAGGCGGCGGCGGCGCATG-dT(30)BN-3',
(sequence: 5'-ATTCTAGGCGGCGGCGGCGCATG-dT(30)BN-3',
where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 282 a 167 c 222 g 211 t
ORIGIN
Alignment Scores:
Pred. No.: 1.18e+04 Length: 882
Score: 293.00 Matches: 67
Percent Similarity: 28.83% Conservative: 12
Best Local Similarity: 24.45% Mismatches: 6
Query Match: 55.08% Indels: 189
DB: 12 Gaps: 56
US-09-854-133-586 (1-97) x BF699490 (1-882)

```



Tue May 13 12:12:51 2003

Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: DCTB/DRP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L12M1276 row: b column: 19  
High quality sequence stop: 706.  
Location/Qualifiers  
1. 910  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4562994"  
/clone\_lib="NIH\_MGC\_14"  
/tissue\_type="renal cell adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 237 a 178 c 249 g 246 t  
ORIGIN  
Alignment Scores: 1.41e+04 Length: 910  
Pred. No.: 292 50 Matches: 55  
Score: 57.66% Conservative: 9  
Percent Similarity: 57.66% Mismatches: 8  
Best Local Similarity: 49.55% Indels: 40  
Query Match: 54.98% Gaps: 21  
DB: 12  
US-09-854-133-586 (1-97) x BG326527 (1-910)  
Qy 6 ArgAspHisAlaSerLeuGlyAsp---Ser-----GluThr-----LeuSerGlnThr 20  
Db 6 AAG---CAC---ACCTTT---GAGTTTTCACCTGTGAACACTATAGCGCTG-----ACA 50  
Qy 21 GluLeuArgLys-----LysGluArgLys-----Lys-----LysArg---Glu 32  
Db 51 GAG-----ACAGTCTGAAGACGAGGAGGAGCATCATCATGATACACACAGACACCA 104  
Qy 33 Arg---LysPheGluAlaAsnCysGlyIleAspPhe-----IleIlePheTrpIle 48  
Db 105 AGTTGAAGATTT-----GTT---TTCCTTCCCTCTGTTTATTT-----140  
Qy 49 Phe---TrpIle---LeuLeuPheSerHisHisTrpIleGlnGluSerLeuLeuCysPro 66  
Db 141 TTCCTCCCTGTGTCTCTACTA-----TGG---TCAGAAAGCCTGTG-TGTCCA 184  
Qy 67 ProSerProLysGluValThrCysArgGluMetLeuThrGlyGlyCysLeuProTrpAla 86  
Db 185 CCATCTCCAAAGAGAGGTACTCTCAGGAAATGTTAAACGGGAGGCTGCTTCTCTGGCA 244  
Qy 87 ThrArgSerHisLeuGlyArgArgLysCysSer 97  
Db 245 ACAAGGAGCCACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 277  
RESULT 10  
BG326527 910 bp mRNA linear EST 27-FEB-2001  
LOCUS BG326527  
DEFINITION 602425373F1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4562994 5',  
mRNA sequence.  
ACCESSION BG326527  
VERSION BG326527  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
1 (bases 1 to 910)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 TITLE NIH-MGC http://mgc.nci.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgsbbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLML)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLML at:  
 http://image.llnl.gov  
 Plate: LLML0460 row: n column: 05  
 High quality sequence stop: 699.

FEATURES  
 source  
 1..937  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:4536772"  
 /tissue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: testis; Vector: pCMV-SPORT6; Site: 1: NotI;  
 Site 2: SalI; Cloned unidirectionally; oligo-dr primed;  
 Average insert size 2.5 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 235 a 202 c 249 g 251 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1-97e+04 Length: 937  
 Score: 290.00 Matches: 69  
 Percent Similarity: 26.4% Conservative: 11  
 Best Local Similarity: 22.8% Mismatches: 11  
 Query Match: 54.5% Indels: 211  
 DB: 12 Gaps: 57

US-09-854-133-586 (1-97) x BG391278 (1-937)

QY 1 GluValGlu-----ValSerArgAspHisAla---SerLeu-----GlyAsp---Ser 14  
 Db 25 GAGGTGGAGGTGCGAGTGGCGGAGATCATGCCACTACATTCAGCTGCCCAATAGAT 84  
 QY 15 GluThrLeuSer-----Gln-----Thr--- 20  
 Db 85 GAGACTGTGTCATAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAGT 144  
 QY 21 -----Glu-----Leu-----Arg----- 23  
 Db 145 TTGGGTTTGGCAGAAATCATCATCAGGAGATATTACAGCCAGCCAGTAGGTAAACCTGAGTG 204  
 QY 24 ---LysLysGlu---Arg---Lys-----Lys-----Lys-----Arg 31  
 Db 205 CTCAGGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 264  
 QY 32 Glu----- 32  
 Db 265 GAGGGGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 324  
 QY 33 -----Arg-----Lys-----Phe 35  
 Db 325 ACACCCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 384  
 QY 36 -----Gln-----AlaAsnCys-----Gly-Ile-----As 42  
 Db 385 ATCTCCAGTGGAGAGATGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 444  
 QY 42 p-----PhelIlePhe-----TrpIle-----Phe---Trp-----IleLeu---Le 53  
 Db 42 p-----Phe-----Trp-----IleLeu---Le 53

Alignment Scores:  
 Pred. No.: 1-97e+04 Length: 937  
 Score: 290.00 Matches: 69  
 Percent Similarity: 26.4% Conservative: 11  
 Best Local Similarity: 22.8% Mismatches: 11  
 Query Match: 54.5% Indels: 211  
 DB: 12 Gaps: 57

US-09-854-133-586 (1-97) x BG391278 (1-937)

QY 1 GluValGlu-----ValSerArgAspHisAla---SerLeu-----GlyAsp---Ser 14  
 Db 25 GAGGTGGAGGTGCGAGTGGCGGAGATCATGCCACTACATTCAGCTGCCCAATAGAT 84  
 QY 15 GluThrLeuSer-----Gln-----Thr--- 20  
 Db 85 GAGACTGTGTCATAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAGT 144  
 QY 21 -----Glu-----Leu-----Arg----- 23  
 Db 145 TTGGGTTTGGCAGAAATCATCATCAGGAGATATTACAGCCAGCCAGTAGGTAAACCTGAGTG 204  
 QY 24 ---LysLysGlu---Arg---Lys-----Lys-----Lys-----Arg 31  
 Db 205 CTCAGGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 264  
 QY 32 Glu----- 32  
 Db 265 GAGGGGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 324  
 QY 33 -----Arg-----Lys-----Phe 35  
 Db 325 ACACCCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 384  
 QY 36 -----Gln-----AlaAsnCys-----Gly-Ile-----As 42  
 Db 385 ATCTCCAGTGGAGAGATGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 444  
 QY 42 p-----PhelIlePhe-----TrpIle-----Phe---Trp-----IleLeu---Le 53  
 Db 42 p-----Phe-----Trp-----IleLeu---Le 53

Db 445 CTGCTGTTT---TTTTTGGTCTTGGTGGCTTTTCTTGGATGGGAATCTGTATTT 501  
 QY 53 u-----PheSer---His-----His-----Trp----- 58  
 Db 502 AAGATGTTAAGTGTGTGTCAGCTTACATATGGAACATCCAGGAGAGTGCACGGCCCC 561  
 QY 59 -----IleGlnGluSer-----Leu-----Leu-----CysProPro---Se 68  
 Db 562 CACCTGCTT---GCAGTGGATTCTGCATCACCAGCTTCAGAGTGT---CCCTGTTC 615  
 QY 68 r-----ProLys-----GluValTh 73  
 Db 616 GCTAGTCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 670  
 QY 73 rCys----- 670  
 Db 671 -TGCTGTGCTAAACGGGAAGAGTCAATCTGTGAAGGTATCATTTTCAAGAAAGAA-- 727  
 QY 77 tLeuThrGly-----Gly-----Cys-----LeuProTrp--- 85  
 Db 728 -TTGGCCGGTGTCCCTTTCACGAGGAGTCCAGTCTCGCGTGGATTTTTCACCTGGGGATG 786  
 QY 86 -----Ala-----Thr-----ArgSer--- 89  
 Db 787 ACGGAACGGGATTATTGGCCGGGGTTCACGGGGTTCGCCATACAAATGGCCGATCTTG 846  
 QY 90 -HisLeuGly-----Arg-----Arg---Lys---Cys----- 96  
 Db 847 GCATTATGGGACGAAAGATACCCCGTTAGAACTCTGTGTTGGGAACAAGAGGACC 906  
 QY 97 -Ser 97  
 Db 907 CTCT 910  
 RESULT 11  
 BE035217  
 LOCUS  
 DEFINITION Mesembryanthemum crystallinum  
 ACCESSION Mesembryanthemum crystallinum cDNA 5', mRNA sequence.  
 VERSION BE035217.1 GI:8330341  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mesembryanthemum crystallinum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Caryophyllales; Caryophyllales; Alismaceae; Mesembryanthemum.  
 Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea  
 ,H., Kawasaki,S., McColough,A., Michalowski,C.B., Palacio,C.,  
 Scarra,G., Wheeler,M. and Zepeda,G.R.  
 Functional Genomics of Plant Stress Tolerance  
 Unpublished (2000)  
 Contact: Michalowski,C.B.  
 University of Arizona  
 Bio Sciences West room 513, Tucson, AZ 85721, USA  
 Tel: 520-621-7982  
 Fax: 520-621-1697  
 Email: chm@u.arizona.edu  
 Best blastx match: 'gil3236235 (AC004684) unknown protein  
 [Arabidopsis thaliana] gi... 116 2e-25'.  
 FEATURES  
 source  
 1..1062  
 /organism="Mesembryanthemum crystallinum"  
 /db\_xref="taxon:3544"  
 /clone\_lib="MO"  
 /tissue\_type="apical meristem and leaf primordia"  
 /dev\_stage="5 weeks"  
 /note="no stress"  
 BASE COUNT 319 a 186 c 245 g 310 t 2 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3.94e+04 Length: 1062

BO688218  
BQ688218.1 GI:21813534  
EST.  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Cloning by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM2367 row: 1 column: 07  
High quality sequence spot: 649.  
Location/Qualifiers  
1. .862  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6209046"  
/clone\_lib="NIH\_MGC\_110"  
/tissue\_type="ductal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pORF7; Site: 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dt priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
Note: this is a NIH\_MGC Library."  
BASE COUNT 158 a 272 c 234 g 198 t  
ORIGIN

Score: 288.50 Matches: 63  
Percent Similarity: 28.62% Conservative: 18  
Best Local Similarity: 22.26% Mismatches: 12  
Query Match: 54.23% Indels: 190  
DB: 10 Gaps: 59  
US-09-854-133-586 (1-97) x BE035217 (1-1062)  
QY 1 Glu-----ValGlu-----Val-----SerArgAspHisAla 9  
Db 103 GAACATTTAAGATCCCTAATGTTGGAACACCTAGTGTCCCGACAGACAAAGCC 162  
QY 10 -----Ser-----LeuGly-----Asp-----Ser-----GluThrLeu----- 17  
Db 163 AAGGACATGTGATTTCTAGATTTGTTTATGAGATATGATGATGACGACTGTCCGAG 222  
QY 18 -----Gln-----ThrGlu-----LeuArgLysLys----- 25  
Db 223 TATTCAGGAGCTCCAGCTAACAGAAATTAATGAACCGTTATGCTGGAGAAAGGTT 282  
QY 25 ----- 25  
Db 283 GAGCAGTACTAGACTAAATGGACAACTCTCTGTTGCAGTTGTTCATCAGAGTTGG 342  
QY 26 Glu-----ArgLysLysLys-----ArgGlu-----ArgLysPhe----- 35  
Db 343 GAGATCTTTACTTACGACGAGAAAGAAACCTTTAAGAAAGATAAGG-----TTTTTTGTAG 399  
QY 36 Gln-----Ala-----Asn----- 38  
Db 400 CAATCAGAGTAGTTTACCTCTCTTTTTCGTTTATGATTTAACAGACACAGTCGCCA 459  
QY 39 -----Cys-GlyIle-----Asp-----Phe-----IleIle----- 45  
Db 460 ATCCATTTACTATTAAATTTGCTAATTTAAATGATTCATTTTGGGAAGTGTATCTGC 519  
QY 46 Phe-----Trp-----IlePhe-----Leu-----Leu-----Phe-----Se 55  
Db 520 GTTTCGATAAAGATGCTCTGTTTATTTGATGTTGGATATGACGTTGATATTCGTC 579  
QY 55 x-----His-----His-----Trp-----IleGln-----GluSerLeu----- 63  
Db 580 GAATGATGATCTTACTTTGATAGATAGATGTTGGGAATCTCAGCGCTGGAGAAATCTCCG 639  
QY 64 -Leu-----Cys-----ProProSer-----ProLysGlu----- 71  
Db 640 CATGCAATACGTGTGATTAAAGAAATACCTCCCTCCGTCAGGAATGATACACCACTCGTG 699  
QY 72 -Val-----Thr-----CysArg-----GluMetLeuThr-----G1 80  
Db 700 AATCTAATCTCATGCTTGTGTNACCTAGAA-----ACCAATAAAGATGGAAGAGAGAGG 753  
QY 80 yGly-----CysLeu-----Pro-----Trp-----Al 86  
Db 754 GGAATCTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 813  
QY 86 a-----Thr-----Arg-----Se 89  
Db 814 TTTTGATGGAACAGAGCTTTAAATTTGAGACCCCGCAGCAATACCTTCGATGATCGG 873  
QY 89 r-----His-----LeuGlyArgLys-----Cys-- 96  
Db 874 ATTTCAATGTTGTCACAAATTAATGATCCCTCCCTGCTCTTGGT-----TATGAATACGCTG 930  
QY 97 -----Ser 97  
Db 931 GATTCG 937

Alignment Scores:  
Pred No.: 1.72e+04 Length: 862  
Score: 287.00 Matches: 62  
Percent Similarity: 32.59% Conservative: 11  
Best Local Similarity: 27.68% Mismatches: 13  
Query Match: 53.95% Indels: 138  
DB: 14 Gaps: 51  
US-09-854-133-586 (1-97) x BQ688218 (1-862)  
QY 1 Glu-----Val-----Glu-----ValSerArg----- 6  
Db 771 GAAGTGTGTAATACTGGACACATAGTTCATGATGCTGAGGAGTC-----AGGAGCGT 715  
QY 7 -----AspHis-----Ala-----SerLeuGlyAspSerGlu-----ThrLeuSer 18  
Db 714 CATGGAGACCATGCTCATTTGGGGTCCAGGAGAGC-----GGGATCCCCAGCTCTTCACG 658  
QY 19 -----GlnThr-----Glu-----Leu-----Arg----- 23  
Db 657 CACTTCAAGGCCCAAGCGGTTATTCGAGACATTTCTCGAAGACGAAATCAAAATC 598  
QY 24 -----Lys-----Lys-----GluArg-----Lys-----LysLysArg 31  
Db 597 TCCAGTGTTTTAGAATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 538  
QY 32 Glu-----Arg-----LysPheGlnAlaAsnCys-----Gly-----IleAspPheIlePheTrp 47  
Db 537 GAGCAAGACAGAGAGGCTCCCACTGTTCCGGGGAGGAGTCTGAA-----TGC----- 490  
QY 48 IlePhe-----TriPheLeuLeu-----PheSerHis----- 56  
|||||

RESULT 12  
BQ688218/c 862 bp mRNA linear EST 15-JUL-2002  
LOCUS AGENCOURT\_8046673 NIH\_MGC\_110 Homo sapiens cDNA clone IMAGE:6209046  
DEFINITION 5', mRNA sequence.

Alignment Scores:

ENCE  
HORS

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
1 (bases 1 to 748)  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,  
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
Russell, D., de Jong, P. and Fraser, C.M.

TITLE  
JOURNAL  
COMMENT

Mouse BAC End Sequences from Library RPCI-24  
Unpublished (1999)  
Other GSSs: RPCI-24-90H13.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 90 row: H column: 13  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. 748  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-90H13"  
/clone\_lib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;  
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The  
library was cloned in the pTARBAC1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."

BASE COUNT 200 a 156 c 176 g 216 t

Alignment Scores:  
Pred. No.: 9.42e+03 Length: 748  
Score: 286.50 Matches: 64  
Percent Similarity: 31.73% Conservative: 15  
Best Local Similarity: 25.70% Mismatches: 11  
Query Match: 53.85% Indels: 159  
DB: 17 Gaps: 54

US-09-854-133-586 (1-97) x AZ740600 (1-748)

Qy 1 GluVal-----Glu-----ValSer-----5  
||||| : : : |||||  
Db 740 GAAGTTACACACCAAGAGTTTCCTTCTGCGCCCAAAATGGCCACTGCCAGGAATAACC 681

Qy 6 Arg-----Asp-----His-----AlaSer-----11  
||| : : : |||  
Db 680 AAGAATGGTTCGGATGAGAGTAACCCACATGTCCTTATAGCCATCATTACCAG 621

Qy 12 GlyAspSer-----GluThrLeu-----Ser-----Glu-----21  
||| ||| ||| ||| ||| : : : ||| : : :  
Db 620 GGG-----TCGGCAGAGCACTGAAGAAATCCAGGAAACAAAGATTATAGTCTTGCAATG 564

Qy 22 Leu-----Arg-----23  
|||  
Db 563 CTGATTACTATAGTTTGGGAGGATATCAACATAAAGGAATGTTCAAAACAACTTGCC 504

Qy 24 -----LysLys-----25  
|||  
Db 503 ATAAATTAACAATAAAGAGCTTATATATTCCTTAAGAAATAATAAGCTTTTAATTA 444

Qy 26 GluArg-----Lys-----LysArg-----GluArgLysPhe 35  
||||| ||| ||| : : : ||| : : : ||| : : :  
Db 443 GAAAGATGAAAGAGGAGCCCTGATATCCCTGACGAGAGAAATATAAGGAGAAGA--- 387

Qy 36 Gln-----AlaAsn-----Cys-----GlyLe-----Asp-----Phe-----43  
||| ||| ||| : : : ||| ||| |||  
Db 386 CAAGATGCACAAATCTGCCGGGAGATAATGTGTCTGAGGACTGCTTTTCCACACAT 327

Qy 44 ---Ile-----IlePheTrpIlePhe---TrpIle---Leu---LeuPhe--- 54  
: : : ||||| ||||| ||| ||| ||| |||||  
Db 326 CAGTTTATTTTCCACACATTTT---ATTTCATAAGTATATGCTTTATTTGTTGT 270

Qy 55 Ser-----HisHisTrpIle---Gln---GluSer---Leu---Leu 64  
||| : : : : : ||| ||| : : :  
Db 269 ACTAGGCTTGAACCTTGAGCTAGGACATTTCTGTAGTAGGACTGAACCTGTGCTAGGCAATG 210

Qy 65 -----Cys-----Pro---Pro-----SerPro-----Lys--- 70  
||| ||| ||| ||| ||| : : :  
Db 209 TGCTCAGTTGATACATCTCCACCCAGCCCTATCCCTAGTCTCCATCTCCAGCT 150

Qy 71 GluValThrCysargGluMetLeuThrGlyGly-----81  
||| ||||| : : : |||||  
Db 149 GAA---ACCTGC-----TTGCTCAGGGGTGGAGCTTCTCTGCTCATGCTTCTGCCACGCC 99

Qy 82 -----Cys-----LeuPro-----TrpAla---ThrArgSerHisLeu-----Gly 92  
||| ||||| ||| ||||| : : : |||||  
Db 98 CACTGCTGGAACCTGCGGCTTGTGTGTGGCCACACACGCTGCTCACCCTTCTACTGGA 39

Qy 93 -----ArgArgLysCysSer 97  
||| ||||| : : : |||  
Db 38 CCTGAGATTATTCGGCGGAA---TCG 15

RESULT 15  
BG261332 1021 bp mRNA linear EST 13-FEB-2001  
LOCUS 602373094F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4484519 5',  
DEFINITION mRNA sequence.  
ACCESSION BG261332  
VERSION BG261332.1 GI:12771148  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [scapbs-r@mail.nih.gov](mailto:scapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10324 row: 1 column: 24  
High quality sequence stop: 689.  
Location/Qualifiers  
1. 1021  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4484519"  
/clone\_lib="NIH\_MGC\_93"  
/tissue\_type="transitional cell papilloma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: bladder; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

BASE COUNT 308 a 220 c 233 g 260 t

ORIGIN

Alignment Scores:  
Pred. No.: 3.87e+04 Length: 1021  
Score: 286.50 Matches: 64  
Percent Similarity: 25.08% Conservative: 15  
Best Local Similarity: 20.32% Mismatches: 15  
Query Match: 53.85% Indels: 221

Search completed: May 11, 2003, 19:06:29  
Job time : 1062 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 16:05:46 ; Search time 360.071 Seconds  
(without alignments)  
1293.204 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FOANGGIDFIIFWIFW 16

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 4101024

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters:  
-MODEL=frame+\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_174132\_704/app.query.fasta\_1.462  
-DB=GenEmbl -QFMT=fastap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=60  
-DOALIGN=200 -THR\_SCORE=QUALITY -THR\_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133.cgn\_1\_1758\_@runat\_05052003\_174132\_704 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rod.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgt\_hum.\*  
40: em\_hgt\_mus.\*  
41: em\_hgtgo\_other.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	16	100.0	337	6	AX321911	Sequence
2	16	100.0	2239	6	AX321909	Sequence
3	16	100.0	5981	6	AX321910	Sequence
4	16	100.0	161280	9	AC093903	Homo sapi
5	8	50.0	171367	9	AL355977	Human DNA
6	8	50.0	176623	2	AC115498	Rattus no
7	7	43.8	579	6	AX390177	Sequence
8	7	43.8	704	8	TOMRNAS	Tomato mRNA
9	7	43.8	720	3	AF082530	Entamoeba
10	7	43.8	730	8	S65047	Sila-glycop
11	7	43.8	777	8	TOMRNASB	Tomato mRNA
12	7	43.8	826	6	E08516	DNA encodi
13	7	43.8	826	6	E12586	cDNA encodi
14	7	43.8	826	6	S65048	self-incomp
15	7	43.8	837	3	AF205061	Dicystoste
16	7	43.8	1209	9	HSBDB12	Homo sapi
17	7	43.8	1283	8	S81597	SI (S11)-se
18	7	43.8	1402	3	SPEATSBP	Sacrophaga
19	7	43.8	1487	1	FNPOPL	Fusobacteri
20	7	43.8	1559	8	AF105149	Zea mays
21	7	43.8	2241	6	E12585	gDNA encodi
22	7	43.8	2241	8	LP028796	Lycoposic
23	7	43.8	3086	8	AF043091	Hordeum v
24	7	43.8	5886	6	AX347114	Sequence
25	7	43.8	5895	3	MTPCOCB	Theileria p
26	7	43.8	8237	1	AWU10505	Acetobacter
27	7	43.8	8576	6	AX347131	Sequence
28	7	43.8	9265	6	AX339183	Sequence
29	7	43.8	9265	6	AX347055	Sequence
30	7	43.8	11046	1	AE001096	Archaeogl
31	7	43.8	11995	1	AE007643	Clostridi
32	7	43.8	12971	1	AE001070	Archaeogl
33	7	43.8	15705	1	AF198256	Haemophil
34	7	43.8	15881	6	AX251764	Sequence
35	7	43.8	15881	6	AX344178	Sequence
36	7	43.8	15881	6	AX345162	Sequence
37	7	43.8	15881	6	AX348575	Sequence
38	7	43.8	16023	2	AC017868	Drosophill
39	7	43.8	16633	6	AX344576	Sequence
40	7	43.8	23933	3	U42843	Caenorhabdi
41	7	43.8	31228	3	U42841	Caenorhabdi
42	7	43.8	34544	3	AF391289	Branchios
43	7	43.8	37977	3	L16621	Caenorhabdi
44	7	43.8	41052	9	AC002499	Human Cos
45	7	43.8	42354	9	AC003032	Human Cos
46	7	43.8	43309	9	AC002498	Human Cos
47	7	43.8	44419	9	AC006506	Human Cos
48	7	43.8	55227	2	AC100331	Homo sapi
49	7	43.8	58518	3	AC024202	Mus muscu
50	7	43.8	58518	3	AC024861	Caenorhab

C 51 7 43.8 65582 9 AL138733 Human DNA  
 C 52 7 43.8 66789 2 AC103807 Homo sapi  
 C 53 7 43.8 68369 2 AC093764 Homo sapi  
 C 54 7 43.8 68837 2 AC091255 Homo sapi  
 C 55 7 43.8 69479 2 AC008597 Homo sapi  
 C 56 7 43.8 78436 2 AC131383 Strongylo  
 C 57 7 43.8 79097 2 AC006897 Caenorhab  
 C 58 7 43.8 83461 3 AE002743 Drosophil  
 C 59 7 43.8 85678 2 AC111726 Rattus no  
 C 60 7 43.8 89936 8 AP004532 Lotu

## ALIGNMENTS

RESULT 1  
 AX321911  
 LOCUS  
 DEFINITION Sequence 442 from Patent WO0172295.  
 ACCESSION AX321911  
 VERSION AX321911.1 GI:17906521  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 REREFERENCE 1  
 AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,  
 Mannion,J. and Kalos,M.D.  
 TITLE Compositions and methods for the therapy and diagnosis of lung  
 cancer  
 JOURNAL Patent: WO 0172295-A 442 04-OCT-2001;  
 CORIXA CORPORATION (US)

FEATURES  
 source  
 Location/Qualifiers  
 1. .337  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 103 a 60 c 93 g 81 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5.69e-10 Length: 337  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX321911 (1-337)  
 QY 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrpIlePheTrp 16  
 Db 107 TTCCAGGCCAATTGGCATAGATTTTATCATATTCTGGATTTTGG 154

RESULT 2  
 AX321909  
 LOCUS  
 DEFINITION Sequence 440 from Patent WO0172295.  
 ACCESSION AX321909  
 VERSION AX321909.1 GI:17906515  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
 AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,  
 Mannion,J. and Kalos,M.D.  
 TITLE Compositions and methods for the therapy and diagnosis of lung  
 cancer  
 JOURNAL Patent: WO 0172295-A 440 04-OCT-2001;  
 CORIXA CORPORATION (US)

FEATURES  
 source  
 Location/Qualifiers  
 1. .2239  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 619 a 444 c 493 g 683 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.08e-09 Length: 2239  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX321909 (1-2239)  
 QY 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrpIlePheTrp 16  
 Db 104 TTCCAGGCCAATTGGCATAGATTTTATCATATTCTGGATTTTGG 151

RESULT 3  
 AX321910  
 LOCUS  
 DEFINITION Sequence 441 from Patent WO0172295.  
 ACCESSION AX321910  
 VERSION AX321910.1 GI:17906518  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
 AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,  
 Mannion,J. and Kalos,M.D.  
 TITLE Compositions and methods for the therapy and diagnosis of lung  
 cancer  
 JOURNAL Patent: WO 0172295-A 441 04-OCT-2001;  
 CORIXA CORPORATION (US)

FEATURES  
 source  
 Location/Qualifiers  
 1. .5981  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 1858 a 1029 c 1098 g 1996 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 7.4e-09 Length: 5981  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX321910 (1-5981)  
 QY 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrpIlePheTrp 16  
 Db 102 TTCCAGGCCAATTGGCATAGATTTTATCATATTCTGGATTTTGG 149

RESULT 4  
 AC093903/c  
 LOCUS  
 DEFINITION Homo sapiens BAC clone RP11-733C7 from 4, complete sequence.  
 ACCESSION AC093903 AC055827  
 VERSION AC093903.3 GI:15920156  
 KEYWORDS  
 SOURCE HTG.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 161280)



AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 2 (bases 1 to 161280)  
 Radionenko, M. and Kozlowicz, A.  
 The sequence of Homo sapiens BAC clone RP11-733C7  
 Unpublished (2001)  
 REFERENCE  
 3 (bases 1 to 161280)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (10-SEP-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE  
 4 (bases 1 to 161280)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (04-OCT-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE  
 5 (bases 1 to 161280)  
 Waterston, R.  
 Direct Submission  
 Submitted (01-MAR-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Oct 4, 2001 this sequence version replaced gi:15625016.  
 -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 -----  
 Summary Statistics  
 Center project name: H\_NH0733C07  
 Drafting Center: WIBR  
 -----

Sulston, J.E. and Waterston, R.  
 Toward a complete human genome sequence  
 Genome Res. 8 (11), 1097-1108 (1998)  
 99063792  
 9847074  
 2 (bases 1 to 161280)  
 Radionenko, M. and Kozlowicz, A.  
 The sequence of Homo sapiens BAC clone RP11-733C7  
 Unpublished (2001)  
 REFERENCE  
 3 (bases 1 to 161280)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (10-SEP-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE  
 4 (bases 1 to 161280)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (04-OCT-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE  
 5 (bases 1 to 161280)  
 Waterston, R.  
 Direct Submission  
 Submitted (01-MAR-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Oct 4, 2001 this sequence version replaced gi:15625016.  
 -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 -----  
 Summary Statistics  
 Center project name: H\_NH0733C07  
 Drafting Center: WIBR  
 -----

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.  
 McPherson, Department of Genetics, Washington University, St. Louis  
 MO. For additional information about the map position of this  
 sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male  
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
 Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
 and coworkers at <http://www.chori.org>  
 VECTOR: pBAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-310A13. Actual start of  
 this clone is at base position 1 of RP11-733C7; actual end is at  
 base position 161280 of RP11-733C7.

Data from AC009792 was used to finish the clone, AC055827.

The sequence of AC055827 has been incorporated into AC093903.

FEATURES	Location/Qualifiers
source	1..161280
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-733C7"
	/clone_lib="RPC1-11"
repeat_region	148..304
	/rpt_family="L1"
repeat_region	306..406
	/rpt_family="L1"
repeat_region	471..525
	/rpt_family="AT-rich"
repeat_region	530..782
	/rpt_family="L1"
repeat_region	685..712
	/rpt_family="(TA)n"
repeat_region	1383..1472
	/rpt_family="GA-rich"
repeat_region	1697..1836
	/rpt_family="L1"
repeat_region	1837..1956
	/rpt_family="L1"
repeat_region	3005..3221
	/rpt_family="MIR"
repeat_region	4013..4321
	/rpt_family="Alu"
repeat_region	4295..4325
	/rpt_family="(A)n"
repeat_region	5625..5960
	/rpt_family="L1"
repeat_region	5938..5982
	/rpt_family="A-rich"
repeat_region	6062..6195
	/rpt_family="MIR"
repeat_region	6388..6462
	/rpt_family="L2"
repeat_region	6546..6868
	/rpt_family="L1"
misc_feature	6866..7318
	/note="similar to EST C06175 (NID:gl502951)"
repeat_region	6866..6898
	/rpt_family="AT-rich"
misc_feature	7134..7385
	/note="similar to EST BG989897 (NID:gl4393967)"
misc_feature	7777..7918
	/note="similar to EST BG388107 (NID:gl3281553)"
repeat_region	8483..8522
	/rpt_family="AT-rich"
repeat_region	8721..9018
	/rpt_family="Alu"
repeat_region	9459..9741
	/rpt_family="L2"
repeat_region	10217..10246
	/rpt_family="AT-rich"
repeat_region	10304..10385
	/rpt_family="MIR"
repeat_region	10835..10878
	/rpt_family="AT-rich"
misc_feature	11861..11987
	/note="similar to EST BG388107 (NID:gl3281553)"
misc_feature	11873..11987
	/note="similar to EST BF286052 (NID:gl1217122)"
repeat_region	12516..12539
	/rpt_family="(T)n"
repeat_region	12929..13232
	/rpt_family="Alu"
repeat_region	13237..13274
	/rpt_family="(TA)n"
repeat_region	13275..13304
	/rpt_family="(T)n"

repeat_region	13696. .13838 /rpt_family="MIR"
repeat_region	14127. .14153 /rpt_family="AT_rich"
repeat_region	14132. .14436 /rpt_family="Alu"
repeat_region	15823. .16117 /rpt_family="Alu"
misc_feature	17329. .17894 /note="similar to EST BG388107 (NID:gl3281553)"
misc_feature	17329. .17618 /note="similar to EST BF286052 (NID:gl11217122)"
misc_feature	17489. .17611 /note="similar to EST AI313891 (NID:g4029010)"
misc_feature	17548. .17869 /note="similar to EST BF095483 (NID:gl10901193)"
misc_feature	17772. .18329 /note="similar to EST BG181259 (NID:gl37303031)"
repeat_region	17845. .17869 /rpt_family="(CTG)n"
repeat_region	18792. .18828 /rpt_family="AT_rich"
repeat_region	20734. .20784 /rpt_family="MERL_type"
repeat_region	20832. .20944 /rpt_family="MERL_type"
repeat_region	20948. .21964 /rpt_family="L1"
repeat_region	21964. .21986 /rpt_family="AT_rich"
repeat_region	21965. .22276 /rpt_family="Alu"
repeat_region	22277. .22323 /rpt_family="L1"
repeat_region	22341. .22681 /rpt_family="L1"
repeat_region	22685. .22824 /rpt_family="Alu"
repeat_region	22805. .22824 /rpt_family="(A)n"
repeat_region	22826. .23575 /rpt_family="L1"
repeat_region	23591. .23735 /rpt_family="MERL_type"

```

Alignment Scores:
Pred. No.: 1.4e-07 161280
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

```

US-09-854-133-587 (1-16) x AC093903 (1-161280)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
 |||||  
 Db 27653 TTCCAGGCCAATTGTGCGATAGATTTTATCATATTTCTGGATTTTGTG 27606

## RESULT 5

AL355977	LOCUS	AL355977	171367 bp	DNA	linear	PRI 09-MAR-2001
	DEFINITION	Human DNA sequence from clone RP11-735G18 on chromosome 10, complete sequence.				

complete sequence.

ACCESSION	AL355977
VERSION	AL355977.11
KEYWORDS	GI:12956926 HTG.

KEYWORDS	SOURCE	ORGANISM
HLG.	human.	<i>Homo sapiens</i>

**ORGANISM**

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 171367)  
LAWLOR, S.

TITLE	COMMENT
<p>Direct Submission</p> <p>Submitted (06-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk</p> <p>On Feb 16, 2001 this sequence version replaced gi:12584715.</p> <p>During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.</p> <p>This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality &gt;= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at</p> <p><a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at</p> <p><a href="http://www.sanger.ac.uk/HGP/Chr10">http://www.sanger.ac.uk/HGP/Chr10</a></p> <p>RP11-735G18 is from the library RPCI-11.3 constructed by the group of Pieter de Jong. For further details see</p> <p><a href="http://www.chori.org/bacpac/home.htm">http://www.chori.org/bacpac/home.htm</a></p> <p>VECTOR: pBACs3.6</p> <p>This sequence is the entire Insert of clone RP11-735G18 The true left end of clone RP11-162E8 is at 131837 in this sequence. The true right end of clone RP11-344N19 is at 88698 in this sequence.</p>	

**FEATURES**  
**source**

us-09-854-133-587.oligo.ige

Tue May 13 12:12:57 2003

```

repeat_region 10001..10127
/note="MIR repeat: matches 20. .147 of consensus"
repeat_region 10152..10480
/note="L2 repeat: matches 2062. .2458 of consensus"
repeat_region 10857..11167
/note="AluJo repeat: matches 5. .310 of consensus"
repeat_region 11367..11443
/note="MIR repeat: matches 77. .154 of consensus"
repeat_region 12813..13122
/note="MLTLD repeat: matches 23. .500 of consensus"
repeat_region 13440..15892
/note="L2 repeat: matches 1630. .2092 of consensus"
repeat_region 16126..16201
/note="L2 repeat: matches 2335. .2412 of consensus"
repeat_region 18939..19041
/note="MIR repeat: matches 158. .262 of consensus"
repeat_region 19709..19791
/note="MIR repeat: matches 1. .80 of consensus"
repeat_region 20433..20476
/note="22 copies 2 mer ga 81% conserved"
repeat_region 20592..20901
/note="L1MB5 repeat: matches 5578. .5887 of consensus"
repeat_region 20912..21222
/note="AluJo repeat: matches 3. .312 of consensus"
repeat_region 21264..21536
/note="AluSc repeat: matches 17. .292 of consensus"
repeat_region 21537..21808
/note="L1MB5 repeat: matches 5896. .6166 of consensus"
repeat_region 23544..23998
/note="L1MB3 repeat: matches 5737. .6176 of consensus"
repeat_region 24311..24394
/note="MIR repeat: matches 43. .140 of consensus"
repeat_region 25430..25854
/note="MIR repeat: matches 1. .426 of consensus"
repeat_region 27916..28009
/note="47 copies 2 mer tt 62% conserved"
repeat_region 28017..28405
/note="LIP repeat: matches 4994. .5386 of consensus"
repeat_region 28591..28862
/note="AluY repeat: matches 40. .307 of consensus"
repeat_region 28892..29188
/note="AluX repeat: matches 3. .297 of consensus"
repeat_region 29270..30294
/note="LIPa7 repeat: matches 5131. .6141 of consensus"
repeat_region 30291..30482
/note="LIPa7 repeat: matches 5055. .5130 of consensus"
repeat_region 30358..30478
/note="L1MC1 repeat: matches 6206. .6325 of consensus"
repeat_region 30479..31518
/note="L1MC1 repeat: matches 4921. .6092 of consensus"
repeat_region 31519..31811
/note="AluY repeat: matches 1. .295 of consensus"
repeat_region 31812..32673
/note="L1MC1 repeat: matches 4078. .4921 of consensus"
repeat_region 32696..33272
/note="MER72 repeat: matches 82. .642 of consensus"
repeat_region 33274..33876
/note="L1MC1 repeat: matches 3473. .4089 of consensus"
repeat_region 33893..34122
/note="MER2 repeat: matches 114. .344 of consensus"
repeat_region 34124..34429
/note="L1MC1 repeat: matches 2966. .3286 of consensus"
repeat_region 34627..34911
/note="AluX repeat: matches 13. .298 of consensus"
repeat_region 35007..35300
/note="MER2 repeat: matches 11. .334 of consensus"
repeat_region 35897..36199
/note="AluX repeat: matches 1. .298 of consensus"
repeat_region 36262..36423
/note="L1MB6 repeat: matches 5798. .5939 of consensus"
repeat_region 36424..36617
/note="LIPa13 repeat: matches 5960. .6151 of consensus"
repeat_region 36618..36787

```

```

Alignment Scores:
Pred. No.: 181
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 50.00%
DB: 9
Length: 171367
Matches: 8
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```

US-09-854-133-587 (1-16) x AL355977 (1-171367)

QY 8 AspPheIlePhePhePhePhe 15  
 |||||  
 Db 36250 GATTTATATATTTTGGATTTT 36273

RESULT 6  
 AC115498 176623 bp DNA linear HTG 13-JUL-2002  
 LOCUS Rattus norvegicus clone CH230-85C21, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 74 unordered pieces.  
 AC115498  
 AC115498.3 GI:21736960  
 VERSION HTG: HTGS PHASE1.  
 KEYWORDS Norway rat.  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 176623)  
 Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,  
 Alsbrooks S.L., Amaratunge H.C., Are J.R., Ayele M., Banks T.,

REFERENCE  
 AUTHORS

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.I., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, I., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, F., Johnson, R., Jolivet, S., Joudah, S., Karcovic, J., Kureshi, A., Landry, N., Leal, B., Korvah, J., Kovar, C., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtarian, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 176623)  
Worley, K.C.  
Direct Submission  
Submitted (20-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 176623)  
Worley, K.C.  
Direct Submission  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced gi:19698616.  
-----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
-----  
Project Information  
Center project name: GNWZ  
Center clone name: CH230-85C21  
-----  
Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 107353 bases at least Q40  
Consensus quality: 114147 bases at least Q30  
Consensus quality: 119101 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 74 contigs. The true order of the pieces

\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1125: contig of 1125 bp in length  
\* 1126: gap of unknown length  
\* 1226: contig of 1616 bp in length  
\* 2841: gap of unknown length  
\* 2842: contig of 1618 bp in length  
\* 4559: gap of unknown length  
\* 4560: contig of 1300 bp in length  
\* 5959: gap of unknown length  
\* 5960: contig of 1052 bp in length  
\* 6060: gap of unknown length  
\* 7111: contig of 1290 bp in length  
\* 7212: gap of unknown length  
\* 8502: contig of 1032 bp in length  
\* 8602: gap of unknown length  
\* 9634: contig of 1360 bp in length  
\* 9734: gap of unknown length  
\* 11094: contig of 1512 bp in length  
\* 11194: gap of unknown length  
\* 12706: contig of 1348 bp in length  
\* 12806: gap of unknown length  
\* 14154: contig of 1094 bp in length  
\* 14254: gap of unknown length  
\* 15347: contig of 1094 bp in length  
\* 15447: gap of unknown length  
\* 15448: contig of 1230 bp in length  
\* 16677: gap of unknown length  
\* 16778: contig of 1473 bp in length  
\* 18251: gap of unknown length  
\* 18350: contig of 1399 bp in length  
\* 19743: gap of unknown length  
\* 19849: contig of 1632 bp in length  
\* 19850: gap of unknown length  
\* 21481: contig of 1156 bp in length  
\* 21482: gap of unknown length  
\* 22737: contig of 1569 bp in length  
\* 22837: gap of unknown length  
\* 22838: contig of 1057 bp in length  
\* 24407: gap of unknown length  
\* 24507: contig of 1166 bp in length  
\* 25563: gap of unknown length  
\* 25564: contig of 1174 bp in length  
\* 25664: gap of unknown length  
\* 26829: contig of 1560 bp in length  
\* 26929: gap of unknown length  
\* 28103: contig of 1163 bp in length  
\* 28203: gap of unknown length  
\* 28204: contig of 1540 bp in length  
\* 28204: gap of unknown length  
\* 28204: contig of 1803 bp in length  
\* 29764: gap of unknown length  
\* 29863: contig of 1409 bp in length  
\* 29864: gap of unknown length  
\* 31026: contig of 1225 bp in length  
\* 31266: gap of unknown length  
\* 31127: contig of 1188 bp in length  
\* 32667: gap of unknown length  
\* 32667: contig of 2312 bp in length  
\* 32767: gap of unknown length  
\* 34570: contig of 1715 bp in length  
\* 34570: gap of unknown length  
\* 36079: contig of 1261 bp in length  
\* 36079: gap of unknown length  
\* 36179: contig of 1514 bp in length  
\* 37404: gap of unknown length  
\* 37504: contig of 1242 bp in length  
\* 38692: gap of unknown length  
\* 38792: contig of 1523 bp in length  
\* 41104: gap of unknown length  
\* 41204: contig of 1970 bp in length  
\* 42918: gap of unknown length  
\* 43019: contig of 1242 bp in length  
\* 43019: gap of unknown length  
\* 44279: contig of 1242 bp in length  
\* 44379: gap of unknown length  
\* 44380: contig of 1242 bp in length  
\* 45894: gap of unknown length  
\* 45894: contig of 1242 bp in length  
\* 47236: gap of unknown length  
\* 47336: contig of 1523 bp in length  
\* 48859: gap of unknown length  
\* 48959: contig of 1970 bp in length  
\* 50928: gap of unknown length

us-09-854-133-587.oligo.rge

Tue May 13 12:12:57 2003

```

* 50929 51028: gap of unknown length
* 51029 contig of 1448 bp in length
* 52477 52576: gap of unknown length
* 52478 contig of 2868 bp in length
* 52577 55444: gap of unknown length
* 55445 55544: gap of 1312 bp in length
* 55545 56956: gap of unknown length
* 56957 56956: contig of 1586 bp in length
* 56958 58542: gap of unknown length
* 58543 58642: contig of 1984 bp in length
* 58643 60626: gap of unknown length
* 60627 60726: contig of 1864 bp in length
* 60727 62590: contig of unknown length
* 62591 62690: gap of unknown length
* 62691 65084: contig of 2394 bp in length
* 65085 65184: gap of unknown length
* 65185 66974: contig of 1790 bp in length
* 66975 67074: gap of unknown length
* 67075 69013: contig of 1939 bp in length
* 69014 69113: gap of unknown length
* 69114 70953: contig of 1840 bp in length
* 70954 71053: gap of unknown length
* 71054 73696: contig of 2843 bp in length
* 73697 73796: gap of unknown length
* 73797 76005: contig of 2209 bp in length
* 76006 76105: gap of unknown length
* 76106 78643: contig of 2538 bp in length
* 78644 78743: gap of unknown length
* 78744 82025: contig of 3282 bp in length
* 82026 82125: gap of unknown length
* 82126 83983: contig of 1858 bp in length
* 83984 84083: gap of unknown length
* 84084 86082: contig of 1999 bp in length
* 86083 86182: gap of unknown length
* 86183 88378: contig of 2196 bp in length
* 88379 88478: gap of unknown length
* 88479 91276: contig of 2798 bp in length
* 91277 91376: gap of unknown length
* 91378 94059: contig of 2683 bp in length
* 94060 94159: gap of unknown length
* 94160 96805: contig of 2646 bp in length

Alignment Scores:
Pred. No.: 185 Length: 176623
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 2 Gaps: 0

US-09-854-133-587 (1-16) x AC115498 (1-176623)
QY 5 CysGlyIleAspPheIlePhe 12
DB 128732 TGTGGCATAGATTTCATTCATTC 128755

RESULT 7
AX390177/c
LOCUS AX390177 579 bp DNA linear PAT 19-MAR-2002
DEFINITION Sequence 5105 from Patent WO0214500.
ACCESSION AX390177
VERSION AX390177.1 GI:19583307
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Escobedo, J., Garcia, P.D., Sudduth-Klinger, J., Reinhard, C.,
Randazzo, F., Lamson, G., Scott, E.M., Zhang, G., Kassam, A., Pot, D. and
Labat, I.
Human genes and gene expression products
Patent: WO 0214500-A 5105 21-FEB-2002.
CHIRON CORPORATION (US) ; Hyseq Inc. (US)

FEATURES
source CDS
Location/Qualifiers
1..579
/organism="Homo sapiens"
/db_xref="taxon:9606" 161 t
BASE COUNT 195 a 123 c 100 g 161 t
ORIGIN
Alignment Scores:
Pred. No.: 15.5 Length: 579
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX390177 (1-579)
QY 10 IleIlePheTrpIlePheTrp 16
DB 328 ATAATATTGGATATTGG 308

RESULT 8
TOMRNAS/c
LOCUS TOMRNAS 704 bp mRNA linear PLN 04-FEB-1999
DEFINITION Tomato mRNA for Slla-RNase, partial sequence.
ACCESSION D17322
VERSION D17322.1 GI:443776
KEYWORDS RNase; ribonuclease; style tissue specific Slla-RNase.
SOURCE Lycopersicon peruvianum mature style cDNA to mRNA, clone I.
ORGANISM Lycopersicon peruvianum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE
1 (bases 1 to 704)
Chung, I., Nakata, K., Tanaka, H., Ito, T., Horiuchi, H., Ohta, A. and
Takagi, M.
Identification of cDNA clones coding for the style specific
Slla-RNase gene associated with gametophytic self-incompatibility
in tomato (Lycopersicon peruvianum)
in Tomato (Lycopersicon peruvianum)
Biosci. Biotechnol. Biochem. (1993) In press
JOURNAL
REFERENCE
2 (sites)
Chung, I.K., Ito, T., Tanaka, H., Ohta, A., Nan, H.G. and Takagi, M.
Molecular diversity of three S-allele cDNAs associated with
gametophytic self-incompatibility in Lycopersicon peruvianum
Plant Mol. Biol. 26 (2), 757-762 (1994)
MEDLINE 95036055
PUBMED 7948929
REFERENCE
3 (bases 1 to 704)
Chung, I.-K.
Direct Submission
Submitted (28-JUL-1993) Il-Kyung Chung, The University of Tokyo,
Dept. of Agricultural Chemistry, Bunkyo-ku, Tokyo 113, Japan
(Tel:03-3812-2111(ex.3085), Fax:03-3812-9246)
On Jan 22, 1994 this sequence version replaced gi:391924.
Submitted (28-JUL-1993) to DDBJ by:
Il-Kyung Chung
Dept. of Agricultural Chemistry
The University of Tokyo
Bunkyo-ku,
Tokyo 113
Japan
Phone: 03-3812-2111 x3085
Fax: 03-3812-9246.
Location/Qualifiers
1..704
/organism="Lycopersicon peruvianum"
/db_xref="taxon:4082"
/tissue_type="mature style"
<1..489
/codon_start=1
/product="Slla-RNase"
/protein_id="BAA04144.1"

FEATURES
source
CDS
Location/Qualifiers
1..704
/organism="Lycopersicon peruvianum"
/db_xref="taxon:4082"
/tissue_type="mature style"
<1..489
/codon_start=1
/product="Slla-RNase"
/protein_id="BAA04144.1"

```

/db\_xref="GI:443777"  
 /translation="HGLWPDIKITILNCPDPAKYAVTGGKFKVKNKHPDILILEA  
 ASLNSQGFWAYQFKKHGTCSDLFNOEKYFDLALILKDKFDLLTFRNKGIIPKSTCT  
 INKIQTIRTVTGVPVNLSCPTMELLEVIGICFNRDASKLIDCDQPKTCTSGNTEIF  
 FG"  
 polyA\_site 235 a 136 c 124 g 209 t  
 BASE COUNT 704

Alignment Scores:  
 Pred. No.: 18.4 Length: 704  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 8 Gaps: 0

US-09-854-133-587 (1-16) x TOMRNASA (1-704)

QY 6 GlyIleAspPheIleIlePhe 12

Db 550 GGATTTGATTTTATTATATT 530

RESULT 9  
 AF082530

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

FEATURES

source

gene

CDS

BASE COUNT

ORIGIN

AF082530 720 bp mRNA linear INV 12-DEC-2000  
 Entamoeba histolytica Grainin 2 (grainin 2) mRNA, complete cds.

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

AF082530 720 bp mRNA linear INV 12-DEC-2000

Alignment Scores:

Pred. No.: 18.8 Length: 720  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x AF082530 (1-720)

QY 1 PheGlnAlaAsnCysGlyIle 7

Db 641 TTTACGGCTAAATGCGGTATT 661

RESULT 10  
 S65047/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REMARK

FEATURES

source

gene

CDS

BASE COUNT

ORIGIN

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

S65047 730 bp mRNA linear PLN 29-OCT-1993

us-09-854-133-587.oligo.rge

Tue May 13 12:12:57 2003

US-09-854-133-587 (1-16) x S65047 (1-730)

QY 6 GlyIleAspPheIleIlePhe 12  
 |||||

Db 550 GGAATGATTTATTTATTT 530

RESULT 11  
 TOMRNASB 777 bp mRNA linear PLN 04-FEB-1999  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

ORGANISM  
 TOMRNASB  
 Tomato mRNA for Slla-RNase, partial sequence.

DI7323.1 GI:443778  
 RNase; ribonuclease; style tissue specific Slla-RNase.  
 Lycopersicon peruvianum mature style cDNA to mRNA, clone II.  
 Lycopersicon peruvianum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.

REFERENCE  
 AUTHORS  
 Chung, I., Nakata, K., Tanaka, H., Ito, T., Horiuchi, H., Ohta, A. and Takagi, M.

TITLE  
 Identification of cDNA clones coding for the style specific Slla-RNase gene associated with gametophytic self-incompatibility in tomato (Lycopersicon peruvianum)

JOURNAL  
 Biosci. Biotechnol. Biochem. (1993) In press  
 2 (bases 1 to 777)

REFERENCE  
 AUTHORS  
 Chung, I.-K.  
 Direct Submission  
 Submitted (28-JUL-1993) Il-Kyung Chung, The University of Tokyo,  
 Dept. of Agricultural Chemistry, Bunkyo-ku, Tokyo 113, Japan  
 (Tel:03-3812-2111(ex.3085), Fax:03-3812-9246)

TITLE  
 On Jan 22, 1994 this sequence version replaced gi:391925.  
 Submitted (28-JUL-1993) to DDBJ by:  
 Il-Kyung Chung  
 Dept. of Agricultural Chemistry  
 The University of Tokyo  
 Bunkyo-ku,  
 Tokyo 113  
 Japan  
 Phone: 03-3812-2111 x3085  
 Fax: 03-3812-9246.

FEATURES  
 source  
 1..777  
 /organism="Lycopersicon peruvianum"  
 /db\_xref="taxon:4082"  
 /tissue\_type="mature style"  
 <1..489  
 /codon\_start=1  
 /product="Slla-RNase"  
 /protein\_id="BA04145.1"  
 /db\_xref="GI:443779"  
 /translation="HGLWPDIKGILNLCNPDANKYASVTGGKFKVKNKHPDILITEA  
 ASLNSQGFWAYQFKKHGTCSDLEFQYFDLALILKDKFLDTFRNKGILPKSTCT  
 INKIQKTIRTVTGVVPLNSCTPTMELLEVGCNFRNDASKLIDCDQPKTDCISGNTLEIF  
 EG"

CDS  
 polyA\_site 777  
 BASE COUNT 260 a 150 c 131 g 236 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 20.1 Length: 777  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 8 Gaps: 0

US-09-854-133-587 (1-16) x TOMRNASB (1-777)

QY 6 GlyIleAspPheIleIlePhe 12  
 |||||

Db 550 GGAATGATTTATTTATTT 530

RESULT 12  
 E08516/c

LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

ORGANISM  
 E08516  
 DNA encoding style-specific S-ribonuclease..

DI7323.1 GI:2176631  
 Lycopersicon peruvianum.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.

REFERENCE  
 AUTHORS  
 Tel, I., Nakada, K., Ito, T., Horiuchi, H., Ota, A., Takagi, M.,  
 Tsubura, H., Tanaka, H. and Ishiguro, Y.

TITLE  
 S-RIBONUCLEASE SPECIFIC TO STYLE AND DNA SEQUENCE CODING THEREFOR

JOURNAL  
 Patent: JP 1994335389-A 1 06-DEC-1994;  
 KAGOME CO LTD

COMMENT  
 OS Lycopersicon peruvianum (tomato)  
 PN JP 1994335389-A/1  
 PD 06-DEC-1994

PF 27-MAY-1993 JP 1993126286  
 PI TEI ITSUIRU, NAKADA KENGO, ITO TORU, HORIUCHI HIROYUKI, PI  
 OTA AKINORI,  
 PI TAKAGI MASAMICHI, TSUBURA HIROKAZU, TANAKA HIROSHI, PI  
 ISHIGURO YUKIO

PC C12N9/22.C12N15/52;  
 CC strandedness: Double;  
 CC topology: Linear;  
 FH Key Location/Qualifiers

FT source 1..826  
 /tissue\_type='pistil'  
 /organism='Lycopersicon peruvianum' FT

FT mat\_peptide 1..486  
 /product='style-specific S-ribonuclease'.  
 FT

FEATURES  
 source  
 1..826  
 /organism="Lycopersicon peruvianum"  
 /db\_xref="taxon:4082"  
 /tissue\_type="pistil"

BASE COUNT 309 a 150 c 131 g 236 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 21.2 Length: 826  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x E08516 (1-826)

QY 6 GlyIleAspPheIleIlePhe 12  
 |||||

Db 550 GGAATGATTTATTTATTT 530

RESULT 13  
 E12586/c

LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

ORGANISM  
 E12586  
 cDNA encoding S-ribonuclease.

E12586  
 E12586.1 GI:3251418  
 JP 1997028381-A/3.

Lycopersicon peruvianum.  
 Lycopersicon peruvianum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.

REFERENCE  
 1 (bases 1 to 826)

QY 6 GlyIleAspPheIleIlePhe 12  
 |||||

Db 550 GGAATGATTTATTTATTT 530





VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	FEATURES	source	gene	CDS	BASE COUNT	ORIGIN	Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:	US-09-854-133-587 (1-16) x AF205061 (1-837)	US-09-854-133-587 (1-16) x S81597 (1-1283)	QY	Db	RESULT 18	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE																		
S81597.1	GI:1478372		Lycopodium peruvianum.	Chung, I.K., Lee, S.Y., Ito, T., Tanaka, H., Nam, H.G. and Takagi, M.	The 5' flanking sequences of two S alleles in Lycopodium peruvianum are highly heterologous but contain short blocks of homologous sequences	Plant Cell Physiol. 36 (8), 1621-1627 (1995)	96146078	8389935		1. 1283	/organism="Lycopodium peruvianum"	/db_xref="taxon:4082"	448..1185	427 a 234 c 204 g 418 t	31.4	7.00	100.00%	100.00%	43.75%	0	0	US-09-854-133-587 (1-16) x S81597 (1-1283)	QY	6	GlycyllelePhe11lePhe 12	Db	1245	GGAATGATTTTATATATTT 1225	RESULT 18	SPEATSBP	1402 bp	mRNA	linear	INV 08-FEB-1999	LOCUS	Sarcophaga peregrina ATBP mRNA for A/T-stretch binding protein	DEFINITION	(transcription factor), complete cds.	ACCESSION	D38160	VERSION	D38160.1	GI:1754528	KEYWORDS	A/T-stretch binding protein; transcription factor; ATBP.	SOURCE	Sarcophaga peregrina cDNA to mRNA.	ORGANISM	Sarcophaga peregrina	REFERENCE	1 (bases 1 to 1402)	Nakanishi-Matsui, M., Kubo, T. and Natori, S.	Molecular cloning and nuclear localization of ATBP, a novel	JOURNAL	Eur. J. Biochem. 230 (2), 396-400 (1995)	MEDLINE	95331271
			Lycopodium peruvianum.	Chung, I.K., Lee, S.Y., Ito, T., Tanaka, H., Nam, H.G. and Takagi, M.	The 5' flanking sequences of two S alleles in Lycopodium peruvianum are highly heterologous but contain short blocks of homologous sequences	Plant Cell Physiol. 36 (8), 1621-1627 (1995)	96146078	8389935		1. 1283	/organism="Lycopodium peruvianum"	/db_xref="taxon:4082"	448..1185	427 a 234 c 204 g 418 t	31.4	7.00	100.00%	100.00%	43.75%	0	0	US-09-854-133-587 (1-16) x S81597 (1-1283)	QY	6	GlycyllelePhe11lePhe 12	Db	1245	GGAATGATTTTATATATTT 1225	RESULT 18	SPEATSBP	1402 bp	mRNA	linear	INV 08-FEB-1999	LOCUS	Sarcophaga peregrina ATBP mRNA for A/T-stretch binding protein	DEFINITION	(transcription factor), complete cds.	ACCESSION	D38160	VERSION	D38160.1	GI:1754528	KEYWORDS	A/T-stretch binding protein; transcription factor; ATBP.	SOURCE	Sarcophaga peregrina cDNA to mRNA.	ORGANISM	Sarcophaga peregrina	REFERENCE	1 (bases 1 to 1402)	Nakanishi-Matsui, M., Kubo, T. and Natori, S.	Molecular cloning and nuclear localization of ATBP, a novel	JOURNAL	Eur. J. Biochem. 230 (2), 396-400 (1995)	MEDLINE	95331271
			Lycopodium peruvianum.	Chung, I.K., Lee, S.Y., Ito, T., Tanaka, H., Nam, H.G. and Takagi, M.	The 5' flanking sequences of two S alleles in Lycopodium peruvianum are highly heterologous but contain short blocks of homologous sequences	Plant Cell Physiol. 36 (8), 1621-1627 (1995)	96146078	8389935		1. 1283	/organism="Lycopodium peruvianum"	/db_xref="taxon:4082"	448..1185	427 a 234 c 204 g 418 t	31.4	7.00	100.00%	100.00%	43.75%	0	0	US-09-854-133-587 (1-16) x S81597 (1-1283)	QY	6	GlycyllelePhe11lePhe 12	Db	1245	GGAATGATTTTATATATTT 1225	RESULT 18	SPEATSBP	1402 bp	mRNA	linear	INV 08-FEB-1999	LOCUS	Sarcophaga peregrina ATBP mRNA for A/T-stretch binding protein	DEFINITION	(transcription factor), complete cds.	ACCESSION	D38160	VERSION	D38160.1	GI:1754528	KEYWORDS	A/T-stretch binding protein; transcription factor; ATBP.	SOURCE	Sarcophaga peregrina cDNA to mRNA.	ORGANISM	Sarcophaga peregrina	REFERENCE	1 (bases 1 to 1402)	Nakanishi-Matsui, M., Kubo, T. and Natori, S.	Molecular cloning and nuclear localization of ATBP, a novel	JOURNAL	Eur. J. Biochem. 230 (2), 396-400 (1995)	MEDLINE	95331271
			Lycopodium peruvianum.	Chung, I.K., Lee, S.Y., Ito, T., Tanaka, H., Nam, H.G. and Takagi, M.	The 5' flanking sequences of two S alleles in Lycopodium peruvianum are highly heterologous but contain short blocks of homologous sequences	Plant Cell Physiol. 36 (8), 1621-1627 (1995)	96146078	8389935		1. 1283	/organism="Lycopodium peruvianum"	/db_xref="taxon:4082"	448..1185	427 a 234 c 204 g 418 t	31.4	7.00	100.00%	100.00%	43.75%	0	0	US-09-854-133-587 (1-16) x S81597 (1-1283)	QY	6	GlycyllelePhe11lePhe 12	Db	1245	GGAATGATTTTATATATTT 1225	RESULT 18	SPEATSBP	1402 bp	mRNA	linear	INV 08-FEB-1999	LOCUS	Sarcophaga peregrina ATBP mRNA for A/T-stretch binding protein	DEFINITION	(transcription factor), complete cds.	ACCESSION	D38160	VERSION	D38160.1	GI:1754528	KEYWORDS	A/T-stretch binding protein; transcription factor; ATBP.	SOURCE	Sarcophaga peregrina cDNA to mRNA.	ORGANISM	Sarcophaga peregrina	REFERENCE	1 (bases 1 to 1402)	Nakanishi-Matsui, M., Kubo, T. and Natori, S.	Molecular cloning and nuclear localization of ATBP, a novel	JOURNAL	Eur. J. Biochem. 230 (2), 396-400 (1995)	MEDLINE	95331271
			Lycopodium peruvianum.	Chung, I.K., Lee, S.Y., Ito, T., Tanaka, H., Nam, H.G. and Takagi, M.	The 5' flanking sequences of two S alleles in Lycopodium peruvianum are highly heterologous but contain short blocks of homologous sequences	Plant Cell Physiol. 36 (8), 1621-1627 (1995)	96146078	8389935		1. 1283	/organism="Lycopodium peruvianum"	/db_xref="taxon:4082"	448..1185	427 a 234 c 204 g 418 t	31.4	7.00	100.00%	100.00%	43.75%	0	0	US-09-854-133-587 (1-16) x S81597 (1-1283)	QY	6	GlycyllelePhe11lePhe 12	Db	1245	GGAATGATTTTATATATTT 1225	RESULT 18	SPEATSBP	1402 bp																										

```

REFERENCE 2 (bases 1 to 1402)
AUTHORS Nakanishi-Matsui,M.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1994) Mayumi Nakanishi-Matsui, Faculty of
Pharmaceutical Sciences, University of Tokyo; 7-3-1, Hongo,
Bunkyo-ku, Tokyo 113, Japan (Tel:03-3812-2111(ex.4820),
Fax:03-5684-2973)
COMMENT On Dec 27, 1996 this sequence version replaced gi:1065611.
FEATURES
    source
        location/Qualifiers
            1..1402
            /organism="Sarcophaga peregrina"
            /db_xref="taxon:7386"
            186..1250
            /codon_start=1
            /product="A/T-stretch binding protein"
            /protein_id="BAA07349.1"
            /db_xref="GI:1065612"
            /translation="MGFPRIISKNKTYTKLGFCLSGDQQQFVWVCHTQEELOTOD
            KEWKHIDQSHNMHGPKEGRTAAQAEAAAMTPLYRKVSENDQQRDDVYS
            TEDEMOKEPKYTEMRAHDDQQTAAVAIDIKLEPSSLQQSAVAQAOQQQQQQQ
            CQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
            MSYAAAAAVASQNVSTMANLIPQELQYKQELQYKQVQVQKHSTNNSTASASSM
            SSDGGERFCFENGGLKFKYHSRLHLHRSVHSKVRFACEICGASFQKSCNLSHR
            KKKHAKGFTTKATLVPSQSF"
    polyA_site 1402
BASE COUNT 507 a 310 c 278 g 307 t
ORIGIN
Alignment Scores:
Pred. No.: 34 Length: 1402
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 3 Gaps: 0
US-09-854-133-587 (1-16) x SPEATSBP (1-1402)
QY 6 GlyIleAspPheIleIlePhe 12
|||||
Db 230 GGTATAGATTATTATTATTTT 210
RESULT 19
FNPMPL/c 1487 bp DNA linear BCT 16-AUG-2000
LOCUS Fusobacterium nucleatum foma gene, strain ATCC 10953.
DEFINITION X72583
ACCESSION X72583.2 GI:9844097
VERSION 40 kDa protein; cell surface protein; foma gene; outer membrane
protein; transmembrane protein.
SOURCE Fusobacterium nucleatum subsp. polymorphum.
ORGANISM Bacteria; Fusobacteria; Fusobacterium.
REFERENCE 1 (bases 1 to 1487)
AUTHORS Bolstad,A.I., Tommassen,J. and Jensen,H.B.
TITLE Sequence variability of the 40-kDa outer membrane proteins of
Fusobacterium nucleatum strains and a model for the topology of the
proteins
JOURNAL Mol. Gen. Genet. 244 (1), 104-110 (1994)
MEDLINE 94316187
PUBMED 8041356
REFERENCE 2 (bases 1 to 1487)
AUTHORS Bolstad,A.I.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-1993) A. I. Bolstad, University of Bergen, Dept
of Biochemistry, Arstadveien 19, 5009 Bergen, NORWAY
REMARK revised by [3] MAT
REFERENCE 3 (bases 1 to 1487)
AUTHORS Bolstad,A.I.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1994) A. I. Bolstad, University of Bergen, Dept
of Biochemistry, Arstadveien 19, 5009 Bergen, NORWAY
REMARK revised by [4]

```

```

REFERENCE 4 (bases 1 to 1487)
AUTHORS Bolstad,A.I.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-1994) A. I. Bolstad, University of Bergen, Dept
of Biochemistry, Arstadveien 19, 5009 Bergen, NORWAY
REMARK revised by [5]
REFERENCE 5 (bases 1 to 1487)
AUTHORS Jensen,H.B.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2000) H.B. Jensen, University of Bergen, Dept of
Molecular Biology, Thormohlensgate 55, 5020 Bergen, NORWAY
COMMENT On Aug 17, 2000 this sequence version replaced gi:551439.
FEATURES
    source
        location/Qualifiers
            1..1487
            /organism="Fusobacterium nucleatum subsp. polymorphum"
            /strain="ATCC 10953"
            /sub_species="polymorphum"
            /db_xref="taxon:76857"
            322..1447
            /gene="foma"
            322..326
            /gene="foma"
            335..1447
            /gene="foma"
            /codon_start=1
            /transl_table=11
            /product="porin"
            /protein_id="CAA51173.2"
            /db_xref="GI:9844098"
            /db_xref="SPTREMBL:Q47905"
            /translation="MKIALVLGLLVGVSAKEVPAPTPAPEKVVYVEKPVIV
            YRDREVAPEWPNQSVQVQYRWYGEVERKKNRDKDNWATKVNAGRIQTLTKVNTF
            EKQTLVETRNHHITNDIDANNKSNGADEYRLRHFNFGKLGSSKVNATSRVEFKQ
            KTNDEKSLGASVLFDAFYIYNNFFKVDKLGRLPGYKYVWKGHGNGEGTPTVHNE
            YHLAFESDFTLPFNALNLEYSNRYREKFEFTDGLKKAEMGELTAVLSNPTPL
            KAGAFELGNAEGDYTNMHQYKRIGEDGTSVDRRDYELYLEPTLQVSYKPTDFVK
            LYAAGADYRNRIITGESEVKRWQPTASACMKYTF"
            335..394
            /gene="foma"
            395..1444
            /gene="foma"
            /product="porin"
            /note="major outer membrane protein, 40kDa"
BASE COUNT 628 a 181 c 254 g 424 t
ORIGIN
Alignment Scores:
Pred. No.: 35.9 Length: 1487
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 1 Gaps: 0
US-09-854-133-587 (1-16) x FNPMPL (1-1487)
QY 9 PheIleIlePheIlePhe 15
|||||
Db 571 TTTATCATCTTTGGATTTT 551
RESULT 20
AF105149/c 1559 bp mRNA linear PLN 02-JUN-1999
LOCUS Zea mays kaurene synthase (KS) mRNA, partial cds.
DEFINITION AF105149
ACCESSION AF105149
VERSION AF105149.1 GI:4959321
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1559)

```

AUTHORS Richman, A.S. and Brandle, J.E.  
 TITLE Partial kaurene synthase gene from Zea mays  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1559)  
 AUTHORS Richman, A.S. and Brandle, J.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-NOV-1998) Southern Crop Protection and Food Research Centre, Agriculture and AgriFood Canada, 1391 Sandford Street, London, ON N5V 4T3, Canada

FEATURES  
 Source  
 1. 1559  
 /organism="Zea mays"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="csul86"  
 /tissue\_type="illuminated leaves; sheaths"  
 /clone\_lib="UMC Clone Distribution Centre"  
 /dev\_stage="5 weeks old"  
 <1..1559  
 /gene="KS"  
 <1..1377  
 /gene="KS"  
 /function="synthesis of (-)-kaurene by ionization  
 dependant cyclization of copalyl diphosphate"  
 /note="terpene cyclase"  
 /codon\_start=1  
 /product="kaurene synthase"  
 /protein\_id="AAD34319.1"  
 /db\_xref="GI:4959322"  
 /translation="GVNVSSDELYHVEASGLHNSLGGYLVNDRLLLEHLKASTVSI  
 EDESILDSIGSRSTLLRQESGGLAKPSLFKEVEHALDGPVFTTLDRLHHRWNIE  
 NFNIEQHWLETPYLSNQHTSDILALSIRDFSSQFTYQOELQHSWKEKRLDQL  
 QFARQLAYFLSAAGTFSPSLDSALRWAKNGVLTIVDFDFVAGSKELENLYM  
 LVEMDEHKVEFYSEQVEIIFSSYISVQNGERASLVQDRSITKHSLVHVEIWLDELK  
 MMTVEWRLSKVPTPEKEYMINASLTGLGPIVLPAFYVGPKISIVYKDPYDELK  
 KLMSCTGLLNDVQTFREYNELGKNSLLVYHGGMSISDAKRLQKPIDTCRRDL  
 LSLVLRSESVPRCKELFWKCKVCYFFYSTDTGFSQVERAKEVDVINEPLKIQ  
 SHTLVSDV"

BASE COUNT 453 a 288 c 362 g 456 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 37.4 Length: 1559  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 8 Gaps: 0

US-09-854-133-587 (1-16) x AF105149 (1-1559)

QY 9 PheillellePheTrpIlePhe 15  
 |||||  
 Db 996 TTCATCATATCTCGATCTTT 976

RESULT 21  
 E12585/c  
 LOCUS 2241 bp DNA linear PAT 27-APR-1998  
 DEFINITION gDNA encoding.  
 ACCESSION E12585  
 VERSION E12585.1 GI:3251417  
 KEYWORDS JP 1997028381-A/2.  
 SOURCE Lycopersicon peruvianum.  
 ORGANISM Lycopersicon peruvianum.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.  
 REFERENCE 1 (bases 1 to 2241)  
 AUTHORS Tei, I., Minami, K. and Takagi, M.  
 TITLE S- RIBONUCLEASE GENE AND PROMOTER SEQUENCE  
 JOURNAL Patent: JP 1997028381-A 2 04-FEB-1997;  
 TEI ITSUKIYON, MINAMI KOUKICHI, TAKAGI MASAMICHI

COMMENT OS Lycopersicon peruvianum  
 PN JP 1997028381-A/2  
 PD 04-FEB-1997  
 PF 24-JUL-1995 JP 1995187557  
 PI TEI ITSUKIYON, MINAMI KOUKICHI, TAKAGI MASAMICHI PC  
 C12N15/09,C07H21/04,C12N1/21//A01H1/00,C12N5/10,C12N9/22, PC  
 (C12N1/21,  
 PC C12N1:19);  
 CC strandedness: Double;  
 CC topology: Linear;  
 CC hypothetical: No;  
 CC anti-sense: No;  
 FH key Location/Qualifiers  
 FH  
 FT source 1..2241 /organism="Lycopersicon peruvianum" FT  
 TATA\_signal 1255..1258  
 FT 5'UTR 1..1365  
 FT exon 1..1596  
 FT intron 1597..1687 /number=1  
 FT exon 1688..2241 /number=2  
 FT CDS join(1366..1596,1688..2101)  
 FT /product="S-ribonuclease"  
 FT 3'UTR 2102..2241.  
 FEATURES Location/Qualifiers  
 source 1..2241 /organism="Lycopersicon peruvianum"  
 /db\_xref="taxon:4082"  
 BASE COUNT 683 a 390 c 444 g 724 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 51.7 Length: 2241  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x E12585 (1-2241)

QY 6 GlyIleAspPheIleIlePhe 12  
 |||||  
 Db 2203 GGAATGATTTTATATATTT 2183

RESULT 22  
 LPU28796/c  
 LOCUS 2241 bp DNA linear PLN 02-OCT-1995  
 DEFINITION Lycopersicon peruvianum self incompatibility (sl2) gene, complete  
 cds  
 ACCESSION U28796  
 VERSION U28796.1 GI:1002595  
 KEYWORDS  
 SOURCE Lycopersicon peruvianum.  
 ORGANISM Lycopersicon peruvianum.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.  
 REFERENCE 1 (bases 1066 to 2241)  
 AUTHORS Chung, I.-K., Leel, S.Y., Masamichi, T. and Nam, H.G.  
 TITLE 5' Flanking sequences of two S alleles in Lycopersicon peruvianum  
 are highly heterologous but contain short blocks of homologous  
 sequences  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2241)  
 AUTHORS Chung, I.-K., Leel, S.Y., Masamichi, T. and Nam, H.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JUN-1995) II-Kyung Chung, Life Science, Pohang  
 University of Science and Technology, Jigok, Pohang, Kyungbuk,

```

FEATURES
  source      Location/Qualifiers
    1. .2241
      /organism="Lycopersicon peruvianum"
      /db_xref="taxon:4082"
      join(1366. .1596,1688. .2101)
      /gene="s12"
      join(1366. .1596,1688. .2101)
      /gene="s12"
      /function="self incompatibility"
      /codon_start=1
      /product="ribonuclease"
      /protein_id="AAA77040.1"
      /db_xref="GI:1002596"
      /translation="MEKTOHTLAFFLLCALPDVYGFNFNOLQVLWRPASFCKKKE
      RPNPNTIHLGPDKIGTLINCNCPDAKYASVTGGFVRKRNKHWPDLILTEAASLSNQ
      GFWAQFKHGTCSDLFNOEKYFDLALIKDFDLITFRNKGIIIPKSTCTNKIQK
      TIRTVGVVPLNSCTPTMELLEVIGICFNDRASKLIDCDQPKCTDTSNTGTEIFP"
BASE COUNT  683 a 390 c 444 g 724 t
ORIGIN
Alignment Scores:
  Pred. No.:      51.7      Length:      2241
  Score:          7.00      Matches:      7
  Percent Similarity: 100.00%      Conservative: 0
  Best Local Similarity: 100.00%      Mismatches: 0
  Query Match:      43.75%      Indels:      0
  DB:              8      Gaps:      0
US-09-854-133-587 (1-16) x LPD28796 (1-2241)
QY  6 GlyIleAspPheIlePhe 12
|||||
Db  2203 GGAATTGATTTTATTATT 2183
RESULT 23
AF043091/c
LOCUS      AF043091      3086 bp      DNA      linear      PLN 05-JAN-1999
DEFINITION Hordeum vulgare dehydrin 6 (dhn6) gene, complete cds.
ACCESSION  AF043091
VERSION     AF043091.1 GI:4105110
KEYWORDS
SOURCE      Hordeum vulgare subsp. vulgare.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 3086)
  AUTHORS   Choi,D.-W., Zhu,B. and Close,T.J.
  TITLE     The barley (Hordeum vulgare L.) dehydrin multigene family:
  sequences, chromosome assignments, and expression characteristics
  of 11 dhn genes of cv. Dicktoo
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 3086)
  AUTHORS   Choi,D.-W., Zhu,B. and Close,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1998) Botany and Plant Science, University of
  California, Riverside, CA 92521, USA
FEATURES
  source      Location/Qualifiers
    1. .3086
      /organism="Hordeum vulgare subsp. vulgare"
      /cultivar="Dicktoo"
      /db_xref="taxon:112509"
      /chromosome="4H"
      670. .>2395
      /gene="dhn6"
      join(670. .1892,1985. .>2395)
      /gene="dhn6"
      /product="dehydrin 6"
      670. .674
      /gene="dhn6"
      join(795. .1892,1985. .2395)
      /gene="dhn6"
gene
mRNA
TATA_signal
CDS

```

```

/note="Lea D-11 protein"
/codon_start=1
/product="dehydrin 6"
/protein_id="AAD02257.1"
/db_xref="GI:4105111"
/translation="MAHFOGQOHHGHPATRVDEYGNPVTAGHGCGGVTGTDLGHFQOQ
QHGRTHDAGYGGSGIAPRHGAGTGVHDAGLGHGTTHGATSTHGTGTAGTGAGVGTGA
GMTGTGHTAGYDAGTGITGHTGHTTGGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
GTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
THDVVTHPHGGLBHKTGCGILHRSGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
GNKEQTATGGYGPYGTGTTGGAGHAGTEGHEKKGVMEKIKELPLGGHKDPPHTT
ATTGYGAGTTCTGCGAGTTCGTGCTGATGTEGHEKKGVMEKIKELPLGGHKDPPHTT"
BASE COUNT  714 a 924 c 939 g 509 t
ORIGIN
Alignment Scores:
  Pred. No.:      68.7      Length:      3086
  Score:          7.00      Matches:      7
  Percent Similarity: 100.00%      Conservative: 0
  Best Local Similarity: 100.00%      Mismatches: 0
  Query Match:      43.75%      Indels:      0
  DB:              8      Gaps:      0
US-09-854-133-587 (1-16) x AF043091 (1-3086)
QY  9 PheIlePheTriPhe 15
|||||
Db  2801 TTTATTATTTTGGATATT 2781
RESULT 24
AX347114
LOCUS      AX347114      5886 bp      DNA      linear      PAT 01-FEB-2002
DEFINITION Sequence 2185 from Patent WO0200928.
ACCESSION  AX347114
VERSION     AX347114.1 GI:18495002
KEYWORDS   .
SOURCE      synthetic construct.
ORGANISM    synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  Patent:  WO 0200928-A 2185 03-JAN-2002;
  Epigenomics AG (DE)
FEATURES
  source      Location/Qualifiers
    1. .5886
      /organism="synthetic construct"
      /db_xref="taxon:32630"
      /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT  1482 a 47 c 1292 g 3065 t
ORIGIN
Alignment Scores:
  Pred. No.:      122      Length:      5886
  Score:          7.00      Matches:      7
  Percent Similarity: 100.00%      Conservative: 0
  Best Local Similarity: 100.00%      Mismatches: 0
  Query Match:      43.75%      Indels:      0
  DB:              6      Gaps:      0
US-09-854-133-587 (1-16) x AX347114 (1-5886)
QY  6 GlyIleAspPheIlePhe 12
|||||
Db  4386 GGGATTGATTTTATAATTTT 4406
RESULT 25
MITPCOB
LOCUS      MITPCOB
DEFINITION Theileria parva mitochondrion DNA.
ACCESSION  Z23263

```

VERSION  
KEYWORDS

Z23263.1 GI:437862  
cytochrome b; cytochrome oxidase; cytochrome oxidase subunit I;  
cytochrome oxidase subunit III; large subunit ribosomal RNA;  
ribosomal RNA.

SOURCE  
ORGANISM

Theileria parva.  
Mitochondrion Theileria parva  
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileridae;  
Theileria.

## REFERENCE

1 (bases 1 to 5895)

Kairo, A., Fairlamb, A.H., Gobright, E. and Nene, V.  
A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA  
sequences and open reading frames for mitochondrially encoded

proteins

EMBO J. 13 (4), 898-905 (1994)

94155854

PUBMED

8112303

## REFERENCE

2 (bases 1 to 5895)

Nene, M.

## JOURNAL

Direct Submission

Submitted (06-JUL-1993) Nene V. M., International Laboratory for

Research on Animal Diseases Nairobi Kenya

## FEATURES

source

1. 5895

/organism="Theileria parva"

/organelle="mitochondrion"

/isolate="Muguga"

/db\_xref="taxon:5875"

/germline

1. 60

repeat\_region

/partial

/standard\_name="inverted terminal repeat"

/citation=[1]

/evidence=experimental

/rpt\_type=TERMINAL

38. 1489

/gene="coi"

38. 1489

/partial

/gene="coi"

/function="electron transport"

/note="The coordinates given is where an open reading

frame starts. This ORF starts at the codon AGT."

/citation=[1]

/codon\_start=1

/evidence=experimental

/transl\_table=4

/product="cytochrome oxidase subunit I"

/protein\_id="CAA80798.1"

/db\_xref="GI:437863"

/db\_xref="SPTREMBL:Q36097"

/translation="SFIFEGLFVFNVSNGNKKIIGISYLVLAYWFGMIGFVMSVLI

RTELSMSGKLTMDTEIYNLFTLHGLIMVFNFMGLFGIGNLYLPVLLGCDV

VPRVNYLSLFPQIGFVLVSSYILEIGSGGTWLTPLPLSLNSVNGIDFIIFGLLA

AGIATLSVNFITFTSVKTIQFVIDRISPAWSIVLTSFLLLSLPVTAVFLMF

LDHNTAMFESSNGDPVLYQHLFWFGHPEVYIMLPFGFIISLLSTYTKEMFG

NOTMILANGFVSRLFGTGYVHIRFGSRYFTVTIILALPTGNKIFNWVTL

QVESIKSLGLVFLVFLVNEVIGTGVGLGNAGIDALHDVTYVVGHEFEVLSIG

AIISMICFIIYIQMLLFGIILSNLSSLIAPFMSVLLTFELPMFTGFSPLPRIP

DYPDEMGNFICTQMLMLVLLKAILFIISL"

complement(2301..3068)

/gene="coiII"

complement(2301..3068)

/gene="coiII"

/function="electron transport"

/citation=[1]

/codon\_start=1

/evidence=experimental

/transl\_table=4

/product="cytochrome oxidase subunit III"

/protein\_id="CAA80799.1"

/db\_xref="GI:437864"

/db\_xref="SPTREMBL:Q36098"

/translation="MRNSAQSYLKINIIFETLYFYSTGLDTLEYIDSTYKNFII

rRNA

MYNQVLLYGTTLKYSVGEFFMNSLTIFINSIREIMTSTMTVMVAIFGMFFSEILV  
FSTIWGFFHLRSLNPILLAEINVEAYLQISDVLTGSLYSILLHVRQESANFETDF  
FMEQLLIGFIFLSQNDHEYSLILSYNNYWTLYFFILTLGLHSLHVCAGGIFVLIIQS  
YFEGDGSQRDEEFNAGYVWHFVEMITALTMLLFLA"

3093..3393

/partial

/product="large subunit ribosomal RNA (LSU1)"

/note="Fragmented rRNA sequence; codes for part of domain  
IV and V of LSU rRNA."

/citation=[1]

complement(3404..3514)

/partial

/product="large subunit ribosomal RNA (LSU3)"

/note="Fragmented rRNA gene; codes for part of domain V of  
LSU rRNA."

/citation=[1]

complement(4365..4403)

rRNA

/product="large subunit ribosomal RNA (LSU2)"

/note="Fragmented rRNA gene; codes for part of domain V of  
LSU rRNA"

/citation=[1]

complement(4423..5583)

/gene="cob"

complement(4423..5583)

/gene="cob"

/function="electron transport"

/standard\_name="cob"

/note="Partial cDNA sequenced; extends from nucleotide

5506 to the end of the open reading frame - poly A

addition on cDNA starts at nucleotide 4368."

/citation=[1]

/codon\_start=1

/evidence=experimental

/transl\_table=4

/product="cytochrome b"

/protein\_id="CAA80800.1"

/db\_xref="GI:437865"

/db\_xref="SPTREMBL:Q36099"

/translation="MEYKVKIKLGNRALTYKLTFTMMNNAHIPSILVPKLNLSNW

NYGFIIGILLQLISGLLTFYFVCKEGAFESLSRLVTEQFGVFLYHSVGSVF

YFFEMFIHIKGMYSKMPWSYSGIVILISIVIAFTGVLPDQGSFAGATVIS

NLEWFGKAKVITFGGTPELTKRFFILHFVPKAVIVILLHLYFLHREGSNPL

TLAEAVALLKYLQILFSDVKELIISFMFGQVGIWTLFQADNDNSLSSSTNP

AHIPEWLLFYATKLVKVPYKVGIVAVVVKLLIILVESRSKQSVSTAHHHRVW

TTSVPLVPALFLGCGIRMVINLDLIIGIYVLLSTTFVQKLDSRVRA"

complement(5552..5619)

rRNA

/partial

/product="large subunit ribosomal RNA (LSU5)"

/note="Fragmented rRNA gene; codes for part of domain IV

of LSU rRNA."

/citation=[1]

5691..5770

rRNA

/product="large subunit ribosomal RNA (LSU4)"

/note="Fragmented rRNA gene; codes for part of domain V of  
LSU rRNA."

/citation=[1]

complement(5783..5895)

repeat\_region

/partial

/standard\_name="inverted terminal repeat"

/note="Contains 12 bp direct and inverted sub-repeats."

/citation=[1]

/evidence=experimental

/rpt\_type=TERMINAL

BASE COUNT 2018 a 889 c 879 g 2109 t

ORIGIN

Alignment Scores:

Pred. No.: 122 Length: 5895

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 43.75% Indels: 0
DB: 3 Gaps: 0
US-09-854-133-587 (1-16) x MITPCOCB (1-5895)

QY 6 GlyileAspPheillellePhe 12
|||||
DB 485 GGTATAGATTATATATATT 505

RESULT 26
LOCUS AWU10505
DEFINITION Acetobacterium woodii DSM 1030 BCT 22-NOV-1999
ACCESSION U10505 U63995
VERSION U10505.2 GI:6014710
KEYWORDS
SOURCE Acetobacterium woodii.
ORGANISM Acetobacterium woodii.
REFERENCE 1 (bases 5035 to 8237)
AUTHORS Forster,A., Daniel,R. and Muller,V.
TITLE The Na(+)-translocating ATPase of Acetobacterium woodii is a
F1F0-type enzyme as deduced from the primary structure of its beta,
gamma and epsilon subunits
Biochim. Biophys. Acta 1229 (3), 393-397 (1995)
95267800
774890
2 (bases 1 to 8237)
AUTHORS Rahlfs,S. and Muller,V.
TITLE Sequence of subunit c of the Na(+)-translocating F1F0 ATPase of
Acetobacterium woodii: proposal for determinants of Na+ specificity
as revealed by sequence comparisons
FEBS Lett. 404 (2-3), 269-271 (1997)
97227959
9119076
3 (bases 1 to 8237)
AUTHORS Rahlfs,S. and Muller,V.
TITLE Sequence of subunit a of the Na(+)-translocating F1F0-ATPase of
Acetobacterium woodii: proposal for residues involved in Na+
binding
FEBS Lett. 453 (1-2), 35-40 (1999)
99330164
10403370
4 (bases 1 to 8237)
AUTHORS Rahlfs,S., Aufurth,S. and Muller,V.
TITLE The Na(+)-F(1)F(0)-ATPase operon from Acetobacterium woodii. Operon
structure and presence of multiple copies of atpE which encode
proteolipids of 8- and 18-kda
J. Biol. Chem. 274 (48), 33999-34004 (1999)
20036535
10567365
5 (bases 5035 to 8237)
AUTHORS Muller,V.
TITLE Direct Submission
Submitted (10-JUN-1994) Institut fuer Mikrobiologie, Goettingen,
37077, Germany
6 (bases 1 to 8237)
AUTHORS Muller,V. and Rahlfs,S.
TITLE Direct Submission
Submitted (07-OCT-1999) Lehrstuhl fuer Mikrobiologie, University of
Munich, Maria-Ward-Str. 1a, Munich 80638, Germany
REMARK Sequence update by submitter
COMMENT On or before Oct 7, 1999 this sequence version replaced gi:4713917,
gi:501146.
FEATURES
source
1..8237
/organism="Acetobacterium woodii"
/strain="DSM 1030"
/db_xref="taxon:33952"
1..8237
/gene="Na+ translocating F1F0 ATP synthase operon"

/note="The ATPase of Acetobacterium woodii is of special
interest because it uses
sodium ions instead of protons as physiological coupling
ion and it
contains multiple genes encoding 8- and 16-kDa
proteolipids."
257..658
/gene="atpI"
257..658
/gene="atpI"
/codon_start=1
/transl_table=11
/product="UncI homolog"
/protein_id="AAF01472.1"
/db_xref="GI:6014712"
/translation="MKTMKIFNKLSLETKIMDSGLISVGLALFSLFQNTPLTALGV
LFGLYSILNFKLMOLTDFDKAMKPSARAQRYVOTRYFLRYLTGVVIVYVAINMPWVN
IIGVLGLVAVKISVLENTTLTKKNASINEA"
724..1386
/gene="atpB"
724..1386
/gene="atpB"
/codon_start=1
/transl_table=11
/product="F1F0 ATPase a subunit"
/protein_id="AAF01473.1"
/db_xref="GI:6014713"
/translation="MEGPKLYGELEFIPITOTLVNTWIMLILVVCVILTKNMQIKR
PGKAQVIAETIVTALDILVQGTMGKMKFAPYMLALFELAFSTAGIYGRSPTAD
LNTVACFDVDFDAVLWFKIKRISQGFPEMPFLFPIINIGEFANPVSLSRFLF
GMDGGIIMALLYSALAGLIFPVPFHYFDIFSGLLQSFIFTMLTVMVFISGA
NM"
1410..1958
/gene="atpE1"
1410..1958
/gene="atpE1"
/codon_start=1
/transl_table=11
/product="F1F0 ATPase c1 subunit"
/protein_id="AAF01474.1"
/db_xref="GI:6014714"
/translation="MIGDMNIVDFIQFLSQFDPVDVVKGFSAIGLIGLAWVAGVGPGI
GQGAAGKAEAVGKNPTKSDIVMILGAAVAETSGIFSLVIALILLFANPITSST
ASVWILSASASAGSIAMIAGIGPGTGGYAGKGAEGAVIRPEKMSAILRVMLLQAV
AQTTGIYALIVALLIMYANPFL"
2001..2249
/gene="atpE2"
2001..2249
/gene="atpE2"
/codon_start=1
/transl_table=11
/product="F1F0 ATPase c2 subunit"
/protein_id="AAC45088.2"
/db_xref="GI:4713918"
/translation="MEGLDFIKACSAIGAGIAMIAGVGPIGIGQGAAGKGAEGVGRQP
EAQSDIIRTMILGAAVAETTGIVGLIVALLIFANPFF"
2306..2554
/gene="atpE3"
2306..2554
/gene="atpE3"
/codon_start=1
/transl_table=11
/product="F1F0 ATPase c3 subunit"
/protein_id="AAF01475.1"
/db_xref="GI:6014715"
/translation="MEGLDFIKACSAIGAGIAMIAGVGPIGIGQGAAGKGAEGVGRQP
EAQSDIIRTMILGAAVAETTGIVGLIVALLIFANPFF"
2572..3126
/gene="atpF"
2572..3126

```

```

/ gene="atpF"
/ codon_start=1
/ transl_table=11
/ product="F1FO ATPase b subunit"
/ protein_id="AAF01476.1"
/ db_xref="GI:6014716"
/ translation="MKNLNAVYKLERGIYPLIFAYAGLVGIDKLLLIATVNFIIIFLL
LEHFFYGVKDIKARODDVAEIVGATEKNTAAAKLQOEYEGLLSDIRAKREIIRN
LHFFYGVKDIKARODDVAEIVGATEKNTAAAKLQOEYEGLLSDIRAKREIIRN
VETLAQKHEAWLDFIDKGGDAQ"
3123..3665
/ gene="atpH"
3123..3665
/ gene="atpH"
/ codon_start=1
/ transl_table=11
/ product="F1FO ATPase delta subunit"
/ protein_id="AAF01477.1"
/ db_xref="GI:6014717"
/ translation="MSLVASKYARALFDVAVDKQDLDEIFSDPKTATDLSFSEKKFMD
LMLTSLATGERKKGIILSRSLNSQYVKNYLMILMKRNFEDIFDIYEAFLKCNH
KNLVARVLTVPIDETLRALLENLAKRNNKVVILENEIDKSIILGAVVYVGDQIID
GSIKNQLSQMKQNNRLH"
3682..5187
/ gene="atpA"
3682..5187
/ gene="atpA"
/ EC_number="3.6.1.34"
/ codon_start=1
/ transl_table=11
/ product="F1FO ATPase alpha subunit"
/ protein_id="AAA79906.2"
/ db_xref="GI:6014711"
/ translation="MNLREEISQIIKNEIDRYEDKLEVVVGVTVIOVGDVARKVHGL
ENAMAGELLAFNEVYGMVLNEDNVCVLLGDDDDIVEGDIVRCITRIKFLQANH
VGRVYNALGPDVGGPIVTDHRRPVEVKAAGVIERESVNPQIOTGYKAIDSMIPG
RGQRLIIGDRTGTALGFDTIINQKEDVICIYVAIGOKDSTVAOIVGOLENNAM
DITIIVSAGAAQLAPLHNTAPISGVYMAEYFMKPTKRITVHGRPSKHGCLPSLVD
PRPPREYPGDVFYHLKLLERAKLKGSSITAPLIETQAGDVSAYIPTNVISI
TGQIFLEAELEFRSIRPAVNPGISVRSVGSQAQISMKKRVAGPLRIEYAOYRELASF
AQFGSLDDETKAQLAKGERIVEILKQDQYPMNVEDQVLLIYAATNGFLDIEVKDI
REFEKLIRFAOKKYPEIMTKVKGKDLSDDEVVAFAECIEAYKKVFSKSY"
5212..6114
/ gene="atpG"
5212..6114
/ gene="atpG"
/ EC_number="3.6.1.34"
/ codon_start=1
/ transl_table=11
/ product="F1FO ATPase gamma subunit"
/ protein_id="AAF79907.1"
/ db_xref="GI:501148"
/ translation="MAENVQDIKPRIKSVNMTQITHAMELVASAKLRKSLAEGRR
PYEAMIESIGRIVERKSGNARNIFMDQREKKTAYIITGDKLAGYNVNAKLVEE
HITDKENAVLTVGSRGRDHRNREYHIOGEYLGISERPNFNAKVTAYVMEGFKN
EYDEVYAVTFKYSTIQAQMKLLPLSREELITSGKVTTEETKEKSKMSDRLT
IMTYEPPELLKYLINLVSSIVYVGSIMESAAEQGARRTAMESATTNANEMIDGLT
LQNRVRQAPITQIEISVGAED"
6218..7618
/ gene="atpD"
6218..7618
/ gene="atpD"
/ EC_number="3.6.1.34"
/ codon_start=1
/ transl_table=11

Alignment Scores:
Pred. No.: 165 Length: 8237
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 1

US-09-854-133-587 (1-16) x AWU10505 (1-8237)
QY 5 CysGlyIleAspPheIleIle 11
Db 2210 TGTGCATTGATTTATTATT 2230

RESULT 27
AX347131 8576 bp DNA linear PAT 01-FEB-2002
LOCUS Sequence 2202 from Patent WO0200928.
DEFINITION AX347131
ACCESSION AX347131
VERSION AX347131.1 GI:18495019
KEYWORDS synthetic construct.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2202 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
source
1..8576
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2122 a 64 c 1723 g 4667 t
ORIGIN
Alignment Scores:
Pred. No.: 171 Length: 8576
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 6

US-09-854-133-587 (1-16) x AX347131 (1-8576)
QY 6 GlyIleAspPheIleIlePhe 12
Db 5203 GGAATAGATTTTATAATATT 5223

RESULT 28
AX339183 9265 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 50 from Patent WO0176451.
DEFINITION AX339183
ACCESSION AX339183
VERSION AX339183.1 GI:18129260
KEYWORDS synthetic construct.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with metabolism
JOURNAL Patent: WO 0176451-A 50 18-OCT-2001;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
source
1..9265
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2686 a 133 c 1801 g 4645 t
ORIGIN
Alignment Scores:
Pred. No.: 183 Length: 9265
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
```

```

DB:
US-09-854-133-587 (1-16) x AX339183 (1-9265)
QY 9 PheIleIlePheTrpIlePhe 15
|||||
Db 525 TTTATTATTTTGGATATTT 545

RESULT 29
LOCUS AX347055
DEFINITION Sequence 2126 from Patent WO0200928.
ACCESSION AX347055
VERSION AX347055.1 GI:18494943
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2126 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES
source
Location/Qualifiers
1..9265
/organism="synthetic construct"
/db_xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2686 a 133 c 1801 g 4645 t
ORIGIN

Alignment Scores:
Pred. No.: 183 Length: 9265
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX347055 (1-9265)
QY 9 PheIleIlePheTrpIlePhe 15
|||||
Db 525 TTTATTATTTTGGATATTT 545

RESULT 30
LOCUS AE001096
DEFINITION Archaeoglobus fulgidus
ACCESSION AE001096 AE000782
VERSION AE001096.1 GI:2689419
KEYWORDS
SOURCE
ORGANISM
Archaeoglobus fulgidus.
Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
Archaeoglobaceae; Archaeoglobus.
1 (bases 1 to 11046)
Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,
Richardson, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,
Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,
Woese, C.R. and Venter, J.C.
The complete genome sequence of the hyperthermophilic,
sulphate-reducing archaeon Archaeoglobus fulgidus
Nature 390 (6558), 364-370 (1997)
98049343
PUBMED 9389475

DB:
US-09-854-133-587 (1-16) x AX339183 (1-9265)
QY 9 PheIleIlePheTrpIlePhe 15
|||||
Db 525 TTTATTATTTTGGATATTT 545

RESULT 30
LOCUS AE001096
DEFINITION Archaeoglobus fulgidus
ACCESSION AE001096 AE000782
VERSION AE001096.1 GI:2689419
KEYWORDS
SOURCE
ORGANISM
Archaeoglobus fulgidus.
Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
Archaeoglobaceae; Archaeoglobus.
1 (bases 1 to 11046)
Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,
Richardson, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,
Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,
Woese, C.R. and Venter, J.C.
The complete genome sequence of the hyperthermophilic,
sulphate-reducing archaeon Archaeoglobus fulgidus
Nature 390 (6558), 364-370 (1997)
98049343
PUBMED 9389475

```

REFERENCE  
AUTHORS

2 (bases 1 to 11046)

Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,  
Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,  
Richardson, D.L., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,  
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,  
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,  
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,  
Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P.,  
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,  
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,  
Woese, C.R. and Venter, J.C.

## Direct Submission

Submitted (15-DEC-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
In order to show the genes in ascending order on the genome, the  
origin of this version has been moved by TIGR to position 2093570  
of the original version and the opposite strand is shown from the  
original version.

On Dec 16, 1997 this sequence version replaced gi:2650492.

FEATURES  
Source

1..11046

/organism="Archaeoglobus fulgidus"

/db\_xref="taxon:2234"

91..876

/gene="AF0140"

91..876

/note="AF0140"

/note="similar to GB:M87049 SP:P27851 PTD:148231 GB:U00096

PID:2367307 percent identity: 31.00; identified by

sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="ubiquinone/menaquinone biosynthesis

methyltransferase (ubiE)"

/protein\_id="AAB91083.1"

/db\_xref="GI:2650497"

/translation="MITVLILRCIMVKSTTRKSADRRYRNRIARFDFETPMELM

NFRKRWKMLFDLIRAEKGLVLEIGATGKNIHPYNNVALDISEKMLKRAVRRAKES

GKVDYFLADANLFPNSDVFVTTFVFCSDVDVGLRFAFRVYKKGRAFYFLEH

MLPKSRELFHPLNLLNPIFFRAFGPEINRDTENIORAGFQIVKRYMLFGSVFLIVAE

KPLDSVDVDEAGSYSLKRSSESSDSSNFRFGSLLPLRLTQLLR"

883..1182

/gene="AF0141"

883..1182

/gene="AF0141"

/note="hypothetical protein; identified by GeneMark;

putative"

/codon\_start=1

/transl\_table=11

/product="A. fulgidus predicted coding region AF0141"

/protein\_id="AAB91091.1"

/db\_xref="GI:2650507"

/translation="MNWKKVGIIVLALAIAGLYVAAMPNGMGVGPWMGYTPMMGYGG

GMMGMMGGYAPQYQSGOGYPAINTDTPQPGIYIPYGYGGHCPMMGYW"

1200..1661

/gene="AF0142"

1200..1661

/note="AF0142"

/note="similar to GB:L09121 PID:155085 SP:P98052 percent

identity: 38.04; identified by sequence similarity;

putative"

/codon\_start=1

/transl\_table=11

/product="Cytochrome C oxidase, subunit II, putative"

/protein\_id="AAB91082.1"

/db\_xref="GI:2650496"

/translation="MRKVLTPFLILIPALISQCNLLSPQAGNPKVYIEAFQFG

YRVVAIOEGDKLYPSPSLKINAGDTVFHVTKDVTGHGTIDGTGIAELLVIP

GRTLVGPTFTDVGKTRVRCNSAGSLHPPMSIDIIVEKIPYVLSALL"

1661..3070

/gene="AF0143"

1661..3070



/gene="AF0143"  
/note="similar to GB:L77117 PID:1592228 percent identity:  
25.53; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAB91081.1"  
/db\_xref="GI:2650499"  
/translation="MVGFAALLYDGGGGEIRFARLLKPRWIFLLTAINLSVFLLI  
IFGLGSAVGSNSFIATVILWALFVEMILFAARVWCAVCPILPVIGEWLKRRL  
YVPEFSSFFPARYNMWIPAGIFLLSLFIFWVTKPLVTALLFSLIATATAMYL  
IYPPYFCRVCPASGYGYHNASLLALRSDEKRCRCEKCLGSEEGYACPMKL  
YPGNENITCGCLCELCRTCDYDLRFLRPFELSRAVRAGKALKLDEAAMGFI  
RFLGAIYEFLFPGSLYWLKDFGNMGPYGVNLENAYLLLPSEFGIVKWLWAGIIVG  
TIAAFGLYVIFSWIGARIAKVDTKQFVSLYSAPYGLLVWAGFVGLLAVIHYI  
FSFLDFEGVEKNAPFVNVKLLIIQALFIAGLASAIVATKRVAKRLTDKPEALTA  
VVSFPHLIALTLIKIAG"  
3075..3455  
/gene="AF0144"  
3075..3455  
/gene="AF0144"  
/note="similar to GB:L09121 PID:155085 SP:P98052 percent  
identity: 34.18; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome C oxidase, subunit II (cbab)"  
/protein\_id="AAB91084.1"  
/db\_xref="GI:2650498"  
/translation="MKKEIALLLVILLSAIPAVLYDYIESRKEKNPEIVIEAHIT  
ESGFGKPNYIVPKSSKVKLVKADVTHGIFIPSLGINELIPAGKEVVEYAEKGE  
VYPMFCSYCSPYHRFMRIIVV"  
complement(3481..3819)  
/gene="AF0145"  
complement(3481..3819)  
/gene="AF0145"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0145"  
/protein\_id="AAB91090.1"  
/db\_xref="GI:2650505"  
/translation="MCRRCNAMSFAVLTPLTGLIDMKYRDMKVPILIOQRK  
ITGIVGVVYVYVVLHSLTDYSLAALMALVFFAITGEYGGKLVHGARTILCLNLR  
KSSQKSNPN"  
complement(4030..4245)  
/gene="AF0146"  
complement(4030..4245)  
/gene="AF0146"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0146"  
/protein\_id="AAB91092.1"  
/db\_xref="GI:2650506"  
/translation="MIIAIVAVVIFLNLFLTPYGYPMGYGGMGYSERYPPGC  
CGPARESSNGSMITKPTGACQGR"  
complement(4308..4832)  
/gene="AF0147"  
complement(4308..4832)  
/gene="AF0147"  
/note="similar to GB:L77117 SP:Q58204 PID:1592308 percent  
identity: 38.55; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical transmembrane protein"  
/protein\_id="AAB91080.1"  
/db\_xref="GI:2650495"  
/translation="MLKLSQALPELSKFFIAGAYTQLVSTFGLGFFAANLIIAANVHG  
LSAFFDGRGELRRTTFEAGYGFMPSLGSAVTIPVSLKYVEATITPTIDLQELSANPE

ILVKSIVSHFPDPSVYVYSAIFNLAVTAMSELIWTFALKNARNVELKQAAVCAIPTAI  
FGLHQVNALVRLQ"  
5068..5325  
/gene="AF0148"  
5068..5325  
/gene="AF0148"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0148"  
/protein\_id="AAB91089.1"  
/db\_xref="GI:2650504"  
/translation="MEARAGHDEEVMKWLVAICIAVGMVMPAFQPLNLAFLYYG  
HESLPITAAASAYEGIVITATAAATAELVHLLLOQFL"  
5470..5724  
/gene="AF0149"  
5470..5724  
/gene="AF0149"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0149"  
/protein\_id="AAB91088.1"  
/db\_xref="GI:2650503"  
/translation="MVEAYLVSGFMLVIGIKYGRKIYVWKAGLIILAGPAVIFAA  
AWIAFTGTSASQERIGLAKLSVVMGLTAGVLSVYVLSKS"  
Alignment Scores:  
Pred. No.: 214 Length: 11046  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 1 Gaps: 0  
US-09-854-133-587 (1-16) x AE001096 (1-11046)  
QY 6 GlyleAspPheIleIlePhe 12  
|||||  
DB 4971 GGCATCGATTATATATATATTT 4991  
RESULT 31  
AE007643 11995 bp DNA linear BCT 27-JUL-2001  
LOCUS Clostridium acetobutylicum ATCC824 section 131 of 356 of the  
DEFINITION complete genome.  
ACCESSION AE007643 AE001437  
VERSION AE007643.1 GI:15024237  
KEYWORDS  
SOURCE  
ORGANISM  
Clostridium acetobutylicum.  
Clostridium acetobutylicum  
Clostridia; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
Clostridium.  
REFERENCE  
AUTHORS  
1 (bases 1 to 11995)  
Nolling, J., Berton, G., Omelchenko, M.V., Markarova, K.S., Zeng, Q.,  
Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, J.I.,  
Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P.,  
Daly, M.J., Bennett, G.N., Koonin, E.V. and Smith, D.R.  
Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum  
J. Bacteriol. 183 (16), 4823-4838 (2001)  
21359325  
11466286  
REFERENCE  
AUTHORS  
2 (bases 1 to 11995)  
Childress, D., Zeng, Q. and Smith, D.R.  
Direct Submission  
Submitted (24-JUL-2001) GRC Sequencing Center Production,  
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100  
Beaver Street, Waltham, MA 02453-8443, USA  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
JOURNAL  
FEATURES  
source  
1..11995  
Location/Qualifiers

```

/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
90..239
/gene="CAC1307"
CDS
90..239
/gene="CAC1307"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK79278.1"
/db_xref="GI:15024238"
/transl_table="MDYFSEDELIKILADPNVKTVEDVQNIKSLCKICFSRCSKE
NTNI"
698..1291
/gene="CAC1308"
CDS
698..1291
/gene="CAC1308"
/codon_start=1
/transl_table=11
/product="Uncharacterized low-complexity protein, related
to YISX/YIBG B. subtilis"
/protein_id="AAK79279.1"
/db_xref="GI:15024239"
/transl_table="MSYNDNKFELHKKRWFVSGHEGKLCLEEDVFRKFDLSKLF
E
QANLAFCEVDNLNENIDFHTSLSCSTYKNNLRGCDYKTDLSYDFSDAVIKNVR
ESKSCWESIFERNVNLNCLNIDYFDTDFSNATDNDVDLSLTPETLLNGVKLRN
IKGIEAFHSINIGTFFKPIILLKSQEAKKWLENCI"
complement(2210..2704)
/gene="CAC1310"
CDS
complement(2210..2704)
/gene="CAC1310"
/codon_start=1
/transl_table=11
/product="Integrase XerD family protein (similarity only
with C-term. part"
/protein_id="AAK79280.1"
/db_xref="GI:15024240"
/transl_table="MINTTKNNLIYNGVYKMSPKTIESQRSVYISNFIKLLK
LHKOLENSNAKFEYKDNMEDGRYDLMTCWCKYIHPMLYTNKIKKVLKAAKIK
NVPEHDLRHNTATLLOQGTDLKVIQERLGHDKDATTANIYSHVNSKQAAATEKLLN
ILKD"
<3281..>3751
/gene="CAC1311"
/notes="possible gene fragment 'Hypothetical protein, CF-36
family"
4007..4267
/gene="CAC1312"
CDS
4007..4267
/gene="CAC1312"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK79281.1"
/db_xref="GI:15024241"
/transl_table="MAEISISPEDIYDSSWLSYNSVRMETSIIYDLPFIFESQYS
KNGKPTIKQIARVINSPLANLNRRNIFEYKKEIKLPNEL"
4285..4905
/gene="CAC1313"
CDS
4285..4905
/gene="CAC1313"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK79282.1"
/db_xref="GI:15024242"
/transl_table="MNKKQKEINNSIYIEDSTHNSNLIYTOEFVEELFHECYKLGFLQG
TYANKIPKSKSTVDLNMVEKLSVEIGIANKLNLDIIINQNKLTNDNLNLSRR
VNSKTLNKKIDDLSHRVEGIDKRTIVLKSLDYIKEDICNKSIIKIDKEDTDL
SLIRNIGNMDEKSKWKQFLPFIATSLSGVLVAIIMLCIKSIK"
complement(5011..5484)
/gene="CAC1314"
CDS
complement(5011..5484)
/gene="CAC1314"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK79283.1"
/db_xref="GI:15024243"
/transl_table="MKFNKITSLLIAFSIIGTSTLSVSATTTMAKTNKSNTNIOKSLYD
SETNLNPTIDLNGTISWNPVNVAVYTTTISKETFNCTEATPVNGIHDAPTISI
SFGADIYSNYGFGTYFAEVNADKNGSLQORDISVYYDNDSDIYGKLTATFPNS"
complement(5498..5944)
/gene="CAC1315"
CDS
complement(5498..5944)
/gene="CAC1315"
/codon_start=1
/transl_table=11
/product="Peptidoglycan-binding domain containing protein"
/protein_id="AAK79284.1"
/db_xref="GI:15024244"
/transl_table="MKNLKLKALSGIVSTVLISSAFAAIPTVIONHPIPKVSA
PVA
TNLNATKSKMFASRTYGTSGRYMTAQOVIDNGGIIQEGDTGEPVADIQRRLGIYDD
GIFGSATYSVVDFQRIDREFHRNGYSPLALDGIQVGAQTWYLK"
6351..6704
/gene="CAC1316"
CDS
6351..6704
/gene="CAC1316"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/protein_id="AAK79285.1"
/db_xref="GI:15024245"
/transl_table="MKKVGIIISIVFLVLSFQSCAAGVSNALSKSSSGSAGFTLA
VCMIIAGIIAIVSKYSGMTIISIVFAIGGIIGIANVGTVDLQIWSILSIFAAALI
IFHIVKKNELTGRK"
complement(6878..7648)
/gene="CAC1317"
CDS
complement(6878..7648)
/gene="CAC1317"
/codon_start=1
/transl_table=11
/product="Potassium channel subunit"
/protein_id="AAK79286.1"
/db_xref="GI:15024246"
/transl_table="MNTTHNKILYELTSSLLALIASIMLIIEISVKLSNETLYIFDT
IDNIIILITDYFLRYISKDKKFKENIDLLSLIPENSIFQGRLLKLSKLLKF
TKLIKFLKLFMAFSLLRKSHETKVKTNPHVYITTVLITGTGMHLTEGLSL
GNALWSEVITITVGYGDISPFPFGRVIAIIMLIGISLMLTGTSTFTFISKKNG
ISYRNEIIEIKSKLDNDELSTEDINDICSLKTLKN"
8086..8598
/gene="CAC1318"
CDS
8086..8598
/gene="CAC1318"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK79287.1"
/db_xref="GI:15024247"
/transl_table="MKKVKYKNRLLALFSLAIITLTSFVGCSESTSKSKSNATAKTTA
KPKPTAQSVLDLKKESAYMTNTVTADNDSNKLGRPNQYTEKVTWKNRKS
QVDCSIELFKNKDDATARKDYLEKIKSPMLTQYIEQDNVLRLIDGVLTPEQSEY
INIENSIFKN"
8988..9689
/gene="CAC1319"
CDS
8988..9689
/gene="CAC1319"
/codon_start=1
/transl_table=11
/product="Glycerol uptake facilitator protein, GLPF"
/protein_id="AAK79288.1"
/db_xref="GI:15024248"
/transl_table="MTIFFAEVLGTLILLIGDGVVNVVVLKNSKHASGWIIVTMGW
AFAVAIPALIFGSYSNQNPALTIALAVIGKVAWAKVPIYLAGQFIAGFVAVLVEV

```

YVDQKSENKTDKLGCTVPAVRNSLINFLECVIGTEVLFVGLGVAQNLKNGMG  
AFVPEGLIIVGLSGGPGYAINPARDLAPRIAHVLVLPINPKGDSNWSYAWIPIIAP  
VVGGLIAGVCMYMLL

gene

9766..10323

CDS

9766..10323

/gene="CAC1320"  
/note="glpp"  
/codon\_start=1  
/transl\_table=11  
/product="Glycerol-3-phosphate responsive antiterminator  
(mRNA-binding), GLPP"  
/protein\_id="AAK79289.1"  
/db\_xref="GI:15024249"

/translation="MNIRELEENPVIAAVKNOEQLOLAVNSEIQTIVFLFGDIVSIK  
EISLKSNNKIGIHDIDVEGTKEVIRYIKETEPDGIISTSQIYKNAIKHL  
IAVRVFIDTISLNNVKSIISECDIAEVLPGVPIKVLIIAKHNSKPPVAGGLIET  
KEVIEALKYGATCVSTTRKEIWM"

gene

10354..11850

CDS

10354..11850

/note="glpk"  
/gene="CAC1321"  
/codon\_start=1  
/transl\_table=11  
/product="Glycerol kinase, GLPK"

Alignment Scores:

Pred. No.: 231 Length: 11995  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x AE007643 (1-11995)

Qy 6 GlylleAspPhelellePhe 12

Db 8480 GCATAGATTTTATTATTTT 8460

RESULT 32

AE001070/c

LOCUS Archaeoglobus fulgidus section 37 of 172 of the complete genome.

DEFINITION AE001070 AE0000782

ACCESSION AE001070.1 GI:2689393

VERSION

KEYWORDS Archaeoglobus fulgidus.

SOURCE Archaeoglobus fulgidus

ORGANISM Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

Archaeoglobaceae; Archaeoglobus.

REFERENCE 1 (bases 1 to 12971)

Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,  
Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,  
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C.,  
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,  
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,  
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,  
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,  
Utterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P.,  
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,  
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,  
Woese, C.R. and Venter, J.C.

The complete genome sequence of the hyperthermophilic,

sulphate-reducing archaeon Archaeoglobus fulgidus

Nature 390 (6658), 364-370 (1997)

JOURNAL 98049343

MEDLINE 9389475

REFERENCE 2 (bases 1 to 12971)

Klenk, H.P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E.,  
Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,  
Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C.,

Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,  
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,  
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,  
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,  
Utterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P.,  
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,  
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,  
Woese, C.R. and Venter, J.C.

TITLE Direct Submission

JOURNAL

REMARK

Submitted (15-DEC-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
In order to show the genes in ascending order on the genome, the  
origin of this version has been moved by tigr to position 2093570  
of the original version and the opposite strand is shown from the  
original version.

On Dec 16, 1997 this sequence version replaced gi:2650129.

COMMENT

FEATURES

source

gene

CDS

1. 12971  
/organism="Archaeoglobus fulgidus"  
/db\_xref="taxon:2234"  
80..295  
/gene="AF0483"  
80..295  
/note="AF0483"  
/note="similar to GB:L77117 PID:1500013 percent identity:  
32.26; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"

/protein\_id="AAB90752.1"  
/db\_xref="GI:2650141"  
/translation="MKRILVCPVCKSEKVELDAGGTGKYCKNGCVGSFILEWTEG  
EYRMKEKEFERKEDEKSKPKGVRED"  
329..1465  
/gene="AF0484"  
329..1465  
/note="AF0484"  
/note="similar to PID:1212775 SP:P51695 percent identity:  
44.53; identified by sequence similarity; putative"

gene

CDS

1450..2187  
/gene="AF0485"  
1450..2187  
/note="AF0485"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="GTP cyclohydrolase II (riba-1)"  
/protein\_id="AAB90751.1"  
/db\_xref="GI:2650140"

/translation="MEVSSHVSKGKPVMLDSEKSAICIPAEVWSGEILNFMKNCID  
EIRLALTKQILNLNRRFNGNLIPIDNLEKISAEERAKFIRELVSGNADLKY  
PGRIFVEETKMGVLEPGIAEACVDLARMAGFAPSAFAPLMTAGGVAGEDYALKF  
AKEHSMVPRIKMDIEFRKSEKIVERVIEATLPTKPYGFRAVGYKTPGLGEIVLVK  
GRVDEGLVLRHISECLTGDVHFSLRCQDQLENALMKIDREGKGVYIYMRGHEG  
IGLKNMAYKLOEGKTDVDANIELGFPDMSRYGIAAQILMDLKVYKIRLLTNPL  
KIEELKKYGVKIVREPIEVEPCFVNLPYLAKKDKMGHLICFND"

gene

CDS

1450..2187  
/gene="AF0485"  
1450..2187  
/note="AF0485"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="A. fulgidus predicted coding region AF0485"  
/protein\_id="AAB90754.1"  
/db\_xref="GI:2650143"

/translation="MLQRLICECAELSSCAAFCLNGRRHAGVCGVVSSESRVFRVI  
FGRDFEESIKRVDEDSIRNLIEAERIYSIVSVSVSGEELIDEALGRLS  
TLYPFDVSDAFSPQSPFFERLERVLEKELTKVIPPKEKIHGSHSTIIGREGY  
KLILSLAKSPYEVKVPVPGVIEGNATSGGGVKLKLTRSDKGNVRALLIDGSSVQOYV  
VITASSRGEVEEVLKLLSGYVRDLQE"

gene

CDS

2165..2551  
/gene="AF0486"  
2165..2551  
/note="AF0486"  
/note="similar to GB:L77117 SP:Q58075 PID:1591374 percent  
identity: 28.92; identified by sequence similarity;

2165..2551  
/gene="AF0486"  
2165..2551  
/note="AF0486"  
/note="similar to GB:L77117 SP:Q58075 PID:1591374 percent  
identity: 28.92; identified by sequence similarity;

2165..2551  
/gene="AF0486"  
2165..2551  
/note="AF0486"  
/note="similar to GB:L77117 SP:Q58075 PID:1591374 percent  
identity: 28.92; identified by sequence similarity;

gene

CDS

2165..2551  
/gene="AF0486"  
2165..2551  
/note="AF0486"  
/note="similar to GB:L77117 SP:Q58075 PID:1591374 percent  
identity: 28.92; identified by sequence similarity;

2165..2551  
/gene="AF0486"  
2165..2551  
/note="AF0486"  
/note="similar to GB:L77117 SP:Q58075 PID:1591374 percent  
identity: 28.92; identified by sequence similarity;

```

/protein_id="AAB90747.1"
/db_xref="GI:2650136"
/translation="MHLPMQGYDRAITVPSGRFLQVVEAVKRGATAIGKKEK
GVIIADTRVSKLEADTIKIIYKIDHEICATSLVADARVLIDRAIEAQNIRLT
YDEPTIVKELAKKICDFKQOQTOYGGVPRPGVSLIAGVDFPKLTDPSPGALLEYK
ATAIGMGRNAVTEFEKEVYRDDLSPDDAMVLGLVAMGLSIESELVPEINIEVGYVRKVD
RTFEVSPLELKPVYRANERIRELLKK"
gene
6332..7036
/gene="AF0491"
CDS
6332..7036
/gene="AF0491"
/notes="similar to GB:L77117 SP:Q58011 PID:1592297 percent
identity: 57.33; identified by sequence similarity;
putative"
/codon_start=11
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB90746.1"
/db_xref="GI:2650135"
/translation="MVSLLKAVTARLKGEEFVLPDYLRDLKEGKEVNFEDLLA
AEEVFKAQKGERASVDELKIKFTGDDVFEARKIILEGEVQITAEQRLEMAKRRQ
IINFISRNIDPTNAPHPPSRIERALEEAKHIDIFKSVEAQVDIVKALKPILPLK
FMEETAIKIPPHGTGRAISALYNFGVGTREWQRDGSWICVMRIPSGMYGLMDLLG
KVAKGEALTKVLRRIIG"
Alignment Scores:
Pred. No.: 247 Length: 12971
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 1 Gaps: 0
US-09-854-133-587 (1-16) x AE001070 (1-12971)
QY 6 GlytIleAspPheIleIlePhe 12
Db 9702 GCATCGATTATTAATAATTT 9682
RESULT 33
AF198256 15705 bp DNA linear BCT 27-APR-2000
LOCUS
Haemophilus influenzae genetic island 1, complete sequence.
DEFINITION
AF198256
ACCESSION
AF198256.1 GI:6739645
VERSION
KEYWORDS
SOURCE
ORGANISM
Haemophilus influenzae.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
1 (bases 1 to 15705)
Chang,C.C.; Gilsdorf,J.R., Dirlita,V.J. and Marrs,C.F.
Identification and genetic characterization of Haemophilus
influenzae genetic island 1
Infect. Immun. 68 (5), 2630-2637 (2000)
20231799
PUBMED
10768954
2 (bases 1 to 15705)
Chang,C.-C., Marrs,C.F., Gilsdorf,J.R. and Dirlita,V.J.
Direct Submission
Submitted (25-OCT-1999) Epidemiology, University of Michigan, 109
Observatory Street, Ann Arbor, MI 48109, USA
JOURNAL
Location/Qualifiers
1..15705
/organism="Haemophilus influenzae"
/strain="Eagan"
/db_xref="taxon:727"
/map="HI0445-HI0446"
251..1477
/note="integrase"
/codon_start=1
/transl_table=11
/product="CP4-57 integrase-like protein"
FEATURES
source
CDS

```

GTDLLFVGRINQNP LKNNRISIIYGNHSLPKCEDAGSIESTTTTRKGNRNPQCQGLF  
 PKIHSLSFGSGALSYEFAYVRLDRNKALSTNQIRSRFGVWTVTSHPMGDTSKNTK  
 HGNPTMKYIPKNNRNLATFTLSVPVTOGGANV"  
 complement(6261. .6572)  
 /note="ORF5"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAF27353.1"  
 /db\_xref="GI:6739652"  
 /translation="MNEARKPTQFLKVLHRLIFSSIGDGYSWGWT SARNYISELER  
 NHLTKGKVRTTKTADGMGQGYRIETADAEQLKQVIAIKAKGGELTAHEEQAYPRF  
 R"  
 complement(6966. .7946)  
 /note="ORF6"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAF27354.1"  
 /db\_xref="GI:6739653"  
 /translation="MKRGHMEPLLDDYYSLNSQVDFINORTNSAIKENLISYVAIEE  
 KIRFLQIEIKDOLYKIGHNESEGFISDSEIYFKERSTECEDKIGLHKDCSSIS  
 SIYHNNEITLDGDDLLSPFNKAI TEYKGIVLHPELILDFCNOSLQHQGVKNYLI  
 LDEYNVTFEADIFSYFNKMEPKTNEERLYIKHTFVNSFDIKISYDDLKLFPE  
 SENAEQNALKEIAKLTIELDKAENNLQKQISEQNSPILGCVHRNDDLLKIA  
 EVNKGWADYPENVKSNQAOTRDYIIREYGVARTAEIEKIACPINRKKK"  
 complement(8025. .8411)  
 /note="ORF7"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAF27355.1"  
 /db\_xref="GI:6739654"  
 /translation="MKSVLSTFGSFVSALDFELFLAILLFIGFLAIFWPILKWP  
 LAFILCAIVFCYLLTIKIKPKPKLEQVDSWAKQLQRPILQIKQEQBNKPF  
 TGAITHVSDQDKETRLGNITNLNKK"  
 complement(8652. .8936)  
 /note="ORF8"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAF27356.1"  
 /db\_xref="GI:6739655"  
 /translation="MRCQAKONLLLSAVNHYYKKSTAFTFVSLVDDEEPAISEYI  
 ALKCKNAKPEIDSRQNSPNEVLETTIVYTHAHNLEDMKKAERIAKR"  
 complement(8933. .10600)  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="phase D3 terminase-like protein"  
 /protein\_id="AAF27357.1"  
 /db\_xref="GI:6739656"  
 /translation="MEIWHEYAKVQSGEIVACRIKQAVYFDDLNPAYFFDEES  
 VNKEFLASRLCPHYKCHLHRSQPIELSDMQTFLFANLIGPKRTDTGLRKYSAY IOVA  
 KNAKSTLAPLVHWKLGSGQDITAAVSRDQARIVFDDARQMCLLSPPLKRLN  
 QOHLINPKNSMLPLAKHSTIEGTNP SLAIVDEIHLHTNSVYSYSALELGOGAR  
 GLLFATAGSNVISACKQHDYCAQIILEGNQNSFLVILFELDESEIDNPENWIR  
 ANPNIGKSTPYLDNFNTIKKARGIPSEWEMLTKRFNWCQGTTPWLGEENVAQCAR  
 YTESDLLHQDCYGLDLSLTDNLTSCLTYFPQCKVRLTIRHYIPFQLNNVANKNR  
 IYRVNROGLIATSGDCIDYDKIRDIDLKDAENFNIKMGFDVWNAHLRTQLOQA  
 LEVEFPQTYQRFSVPKASAEVLIINQVIEHNGDQVPLSVASHTDANANIKPNA  
 KKAANKIDPAKVFSGTYQYGLDILFELSEEHKQALEQFGNIDL"  
 complement(10607. .10975)  
 /note="ORF9"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAF27358.1"  
 /db\_xref="GI:6739657"

```

/protein_id="AAF27347.1"
/db_xref="GI:6739646"
/translation="MARIVKPLSPTEIKNAKPKKEYTLSDGEGILLILLPSPGSKSWR
FNVAQTKRCKMAGGVPELSIDARAKREYRVLLAKIGDPOEKIRIQOEYENR
LADTPHSVAESYENGIIKYEKANPTREKNWERLKNHIFPYIGDKHVSEIKVKELVSI
YEKIADRSNTAKIHOIYCAIMHATKGTITESHNCRNAVKNVFIKSTPHTIKLDE
LSKLFODLENARIGKATYILLICWSELTALRPEKAVNAEWSIEDFDKNLWNIPEKMG
QADKKRPHVPUSSQAIQLLEVPKLFSENSFVFAGRSSKNQPMKATVNVALKHICY
KOKLTAHTGASPIKTFKTLFAGYVERNVSSETILLSHLEGGDDLENTYNNRYNLEERKPYM
QLIGDYCEGCMNLN"
complement(1649. .3418)
/note="primase"
/codon_start=1
/transl_table=11
/product="phage phi-r73 primase-like protein"
/protein_id="AAF27348.1"
/db_xref="GI:6739647"
/translation="MAKLINAPLADQIHPEYSDIFILAGRAKWAQDNGKGQEWLLL
CSLHGLDRNIKPVIAELKOLENISIRIVKEDQOSVKLYOYQELAGAEITAICQNL
KNSDAIDVLLDAAOAKEDLSAYTQIRLNDEKADLAELQAPPEKAEKQDGTIERL
AFQNLMLDLAGORREIYAVDGTQWKNQDNDLEEKAVKFLDENEFYSDSTIERL
IKTKAQLPMGSENGKRLIAVLVYALITRNINWQWFFETITGKSGSGHVSFASIAITLL
LFDKWLISVSDGKNDKARMLIAVYALITRNINWQWFFETITGKSGSGHVSFASIAITLL
AGVNTASSNLEDFDERGLSENKLTILJCPEQSIGDGGSLKSTGGDTRVRYIN
YQDPFDVKITALVMLNNRPPCSTERSGGVRRRVIFDFKIVPEDERDPHMDKITL
EVGSLIKVDSFDPNDKALKAKQAEQEALEVKLLDPLTDFGYFYTTEQDGL
FIGYVWGLDKITHLYPVLAVYTKAMNIGELGNFVIGVEQALKQNGKNHNMKRIH
TKGRRINIHFKOPDFSENMES"
complement(3447. .3797)
/note="ORF1"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF27349.1"
/db_xref="GI:6739648"
/translation="MSKTAFTAGVHVFGFSGMPVIAMRSKEKAILLEHRTYTLQ
KPTNIRAINLHYTLFELISLGNARAPIHAEGKREAEAEATAPFTLYPTAQIRVIQ
FKDAFLFKQACGLD"
complement(3790. .4428)
/note="ORF2"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF27350.1"
/db_xref="GI:6739649"
/translation="MKPSNPMQOLQWKNNEORKINOFVNMNRNKNLASDYSVSGN
GVHKNKIELVKNCEAKKQIYKVYNNLHSEONTQCNARSEAKSKYQGLKYNALPLAF
KFAQUTROFKLILDSNRKQIYVFDDEPHKLKRDDEMDLIDRLKAGGKLFNALAKSQ
GGTFCRDQOATLKDFNQANGVLIHKFGEVVAQIERLNIERVEGEKILQAGSNE"
complement(4801. .5172)
/note="ORF3"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF27351.1"
/db_xref="GI:6739650"
/translation="MFGOSILSALQVWADFNDINNYKOIPIPMSEKELDLRSESEML
CPSIEVGLMHGLDMQTLNKHMLTQYQFIFALTRAPQKIRILLADTEQOASRRFTD
ASLSEYVLAGGLDGEIERND"
complement(5165. .6094)
/note="ORF4"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF27352.1"
/db_xref="GI:6739651"
/translation="MNSKHSYIQTKNFNITASLOQSAVNAFKVPAQSQKTTASREKLN
YLHQANDSTPPFRAPFFIRTPQSNADVFVNLNPILISVMNEREQSLTGCPVIVAFY
HLVTRPVTSVLAIVLENKHMELTQYQFIFALTRAPQKIRILLADTEQOASRRFTD

```

/translation="MTKTKYKAPDFDEIAKQWAKARITQISYRGDIKPEDLTNLEI  
 YCINYSIFRSADIAKNGFSIVNSQGTQRNPALSAKADAQKVMKMSALLGDFPVS  
 RKNPVETDVIDMDEILTN"  
 complement(111172..11558)  
 /note="holin"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phage phi-105 holin-like protein"  
 /protein\_id="AAF27359.1"  
 /db\_xref="GI:6739658"  
 /translation="MPYQPLRRCSPGCRNKKVAGRCCEHKPKDTRPSSSARGYDHHK  
 SYTAAQYLKHLPLCYMCLPQGYTPATVIDHLKPVENGQADPLFWASNHQPLCRDCH  
 SYKTVIDQRFGEAKDGNPSNLIKQ"  
 complement(11573..11887)  
 /note="ORF10"  
 /codon\_start=1  
 /evidence=not experimental  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAF27360.1"  
 /db\_xref="GI:6739659"  
 /translation="MVEINLIDEIKAHNLHDLDDELLEAYKSTALEVCRKHGKT

Alignment Scores:

Pred. No.: 293 Length: 15705  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x AF198256 (1-15705)

QY 9 PheIlePheTrpIlePhe 15

Db 7380 TTCATTATATTTGGATATTC 7400

RESULT 34  
 AX251764  
 LOCUS  
 DEFINITION  
 AX251764  
 ACCESSION  
 AX251764  
 VERSION  
 AX251764.1 GI:15985119  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 synthetic construct.  
 synthetic construct.  
 artificial sequences.  
 1 (bases 1 to 15881)  
 REFERENCE  
 Olek, A., Piepenbrock, C. and Berlin, K.  
 TITLE  
 Diagnosis of diseases associated with the cell cycle  
 JOURNAL  
 Patent: WO 0168911-A 25 20-SEP-2001;  
 Epigenomics AG (DE)

FEATURES  
 source  
 Location/Qualifiers  
 1..15881  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4515 a 204 c 3588 g 7574 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 296 Length: 15881  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX251764 (1-15881)  
 QY 10 IleIlePheTrpIlePheTrp 16  
 Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 35  
 AX344178

LOCUS  
 DEFINITION  
 AX344178  
 ACCESSION  
 AX344178  
 VERSION  
 AX344178.1 GI:18492066  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 synthetic construct.  
 synthetic construct.  
 artificial sequences.

REFERENCE  
 1  
 AUTHORS  
 Olek, A., Piepenbrock, C. and Berlin, K.  
 TITLE  
 Diagnosis of diseases associated with signal transduction  
 JOURNAL  
 Patent: WO 020926-A 25 03-JAN-2002;  
 Epigenomics AG (DE)

FEATURES  
 source  
 Location/Qualifiers  
 1..15881  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4515 a 204 c 3588 g 7574 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 296 Length: 15881  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX344178 (1-15881)  
 QY 10 IleIlePheTrpIlePheTrp 16  
 Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 36  
 AX345162  
 LOCUS  
 DEFINITION  
 AX345162  
 ACCESSION  
 AX345162  
 VERSION  
 AX345162.1 GI:18493048  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 synthetic construct.  
 synthetic construct.  
 artificial sequences.

REFERENCE  
 1  
 AUTHORS  
 Olek, A., Piepenbrock, C. and Berlin, K.  
 TITLE  
 Diagnosis of diseases associated with the immune system  
 JOURNAL  
 Patent: WO 020928-A 233 03-JAN-2002;  
 Epigenomics AG (DE)

FEATURES  
 source  
 Location/Qualifiers  
 1..15881  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4515 a 204 c 3588 g 7574 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 296 Length: 15881  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX345162 (1-15881)  
 QY 10 IleIlePheTrpIlePheTrp 16  
 Db 11862 ATTATTTTGGATATTTGG 11882

Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 37  
LOCUS AX348575 15881 bp DNA linear PAT 06-FEB-2002  
DEFINITION Sequence 33 from Patent WO0202807.  
ACCESSION AX348575  
VERSION AX348575.1 GI:18614610  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequences.  
REFERENCE  
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of diseases associated with cell signalling  
JOURNAL Patent: WO 0202807-A 33 10-JAN-2002;  
EpiGenomics AG (DE)  
FEATURES  
source  
1..15881  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4515 a 204 c 3588 g 7574 t  
ORIGIN

Alignment Scores:  
Pred. No.: 296 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX348575 (1-15881)

QY 10 IleilePheTrpPhePhe 16  
|||||

Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 38  
LOCUS AC017868/c 16023 bp DNA linear HTG 09-DEC-1999  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AC017868  
VERSION AC017868.1 GI:6553322  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
TITLE Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
JOURNAL Ephydroidea; Drosophilidae; Drosophila.  
COMMENT 1 (bases 1 to 16023)  
Adams, M. and Venter, J. C.  
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
This sequence was identified as CDM:10212420 by the submitter.  
For more information on this record e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1..16023  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 5472 a 3032 c 2604 g 4915 t  
ORIGIN

Alignment Scores:  
Pred. No.: 299 Length: 16023  
Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 2 Gaps: 0

US-09-854-133-587 (1-16) x AC017868 (1-16023)

QY 9 PheileilePheTrpPhePhe 15  
|||||

Db 2500 TTCAATAATTTTGGATATTT 2480

RESULT 39  
LOCUS AX344576 16633 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 1 from Patent WO0200927.  
ACCESSION AX344576  
VERSION AX344576.1 GI:18492462  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequences.  
REFERENCE  
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of diseases associated with development genes  
JOURNAL Patent: WO 0200927-A 1 03-JAN-2002;  
EpiGenomics AG (DE)  
FEATURES  
source  
1..16633  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4452 a 142 c 3667 g 8372 t  
ORIGIN

Alignment Scores:  
Pred. No.: 309 Length: 16633  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX344576 (1-16633)

QY 9 PheileilePheTrpPhePhe 15  
|||||

Db 11972 TTATATTTTGGATTTT 11992

RESULT 40  
LOCUS UA2843/c 23933 bp DNA linear INV 04-JUL-2002  
DEFINITION Caenorhabditis elegans cosmid T19E7, complete sequence.  
ACCESSION UA2843  
VERSION UA2843.2 GI:14574233  
KEYWORDS HTG.  
ORGANISM Caenorhabditis elegans.  
REFERENCE  
AUTHORS Caenorhabditis elegans  
TITLE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
JOURNAL Rhabditoides; Rhabditidae; Peloderinae; Caenorhabditis.  
COMMENT 1 (bases 1 to 23933)  
Waterston, R.  
Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)  
95069613  
9851916  
2 (bases 1 to 23933)  
Pauley, A. and Gattung, S.  
The sequence of C. elegans cosmid T19E7  
Unpublished (2001)  
JOURNAL  
REFERENCE  
AUTHORS 3 (bases 1 to 23933)  
Waterston, R.  
Direct Submission  
TITLE

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (10-DEC-1995) Robert Waterston  
4 (bases 1 to 23933)  
Waterston, R.  
Direct Submission  
Submitted (28-JUN-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
5 (bases 1 to 23933)  
Waterston, R.  
Direct Submission  
Submitted (23-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
6 (bases 1 to 23933)  
Waterston, R.  
Direct Submission  
Submitted (04-JUL-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

On Jun 28, 2001 this sequence version replaced gi:1125791.  
Submitted by:

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
Email: rwhematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate chemistry  
or covered by high quality data (i.e., phred quality >= 30); an  
attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by sequence from  
more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its  
analysis see:

<http://www.wormbase.org/db/seq/sequence?name=tl9e7;class=Sequence>

## NEIGHBORING COSMID INFORMATION

The 5' cosmid is B0547, 1400 bp overlap; the 3' cosmid is C55C3,  
200 bp overlap. Actual start of this cosmid is at base position 401  
of TL9E7; actual end is at 12238 of C55C3.

## NOTES:

Coding sequences below are the result of integration and manual  
review of the following data : computer analysis using the program  
GeneFinder (P. Green and L. Hillier, personal communication), the  
large scale EST projects of Yuji Kohara  
([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C.  
elegans ORFome cloning project (<http://worfdb.dfci.harvard.edu/>),  
similarity to other proteins from BlastX analyses  
(<http://blast.wustl.edu/>), sequence conservation with C. briggsae  
using Jim Kent's WABA alignment program (Genome Research  
10:1115-1125, 2000), individual C. elegans GenBank submissions,  
and personal communications with C. elegans researchers. tRNAs  
are predicted using the program tRNAscan-SE ( Lowe, T.M. and  
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES  
Source

Location/Qualifiers  
1. .23933  
/organism="Caenorhabditis elegans"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="IV"  
/clone="TL9E7"  
complement(36..8656)

## gene

/gene="skn-1"  
/note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=tl9e7.2a;class=Sequence>"  
complement(join(36..112,165..532,686..944,1621..1773,  
1827..2148,3455..3879,4042..4180,8528..8656))  
/gene="skn-1"  
/note="coded for by the following C. elegans cDNAs:  
yk287d4.3, yk18ell.5, yk155h9.5, CESAB55F, yk194d1.5,  
yk202g2.5, yk250e4.5, yk287d4.5, yk314h8.5, yk320b11.5,  
yk385h2.3, yk403d12.5, yk415e5.5, yk421f9.5, yk204f2.5,  
yk251g1.5, yk291h2.5, yk311c1.5, yk2d12.3, CEMSE62F,  
CESAC19P, yk250e4.3, yk415e5.3, yk251g1.3, yk291h2.3"  
/codon\_start=1  
/product="C. elegans SKN-1 protein (corresponding sequence  
TL9E7.2a)"  
/protein\_id="AAAB3594.2"  
/db\_xref="GI:14574234"  
/translation="MGSSRRQSTSATRRDQRRRCQFSSVADDEETTSIYGVSS  
IFWILATSSILVLISSPSSNTSQSSYDRITTKHLLDNISPTFKMYTDSNNRPFDE  
VNHQHQEQDFNGSKIDYQPNRPMGLURRDDQRMETFMNSGVTVPVMPILTEH  
PPASPPGRGPESTERTSSRYEYSPSLEDIDLIDVLRSDIAGEKGTQVAPADQYE  
CDLQTLTEKSTVAPLTAEANARYEDLSKGFYNGFESFNNOYQKHQOQREQIKTP  
TLEHPTQAELEDDLFDAQLFEDVSRREGQLNQLFDNKKQHPVNNVLSSEGLVY  
NQANLTEMQMRDSCNOVISITPTSTAQPTETLFTNVTDSQTEVQWLPEVVPNDVFP  
TSNAYIGMONDSLQAVVNSGQIDYHSYQSTGTPTPLIIGSGRQQTOTSPGV  
TVTATATQSLFDPYHSRHSFDCCTTDSSTCSRLSSSPRYTSSESTGTHESRYGK  
LAISSGRIQRSPSRSSQSIKTARVVPVPLASGQRGRQKDEQLASDNLPLVSAFQ  
ISMSSELUQVLKNESEYQRLIRKRRGNKVAARTCRQRTDRHDKMSHYI"  
complement(join(36..112,165..532,686..944,1621..1773,  
1827..2148,3455..3877))  
/gene="skn-1"  
/note="coded for by the following C. elegans cDNAs:  
yk782b12.5, yk782b12.3"  
/codon\_start=1  
/product="C. elegans SKN-1 protein (corresponding sequence  
TL9E7.2c)"  
/protein\_id="AAM75366.1"  
/db\_xref="GI:21693910"  
/translation="MYTDSNNRNFDEVNHQHQEQDNGQSKYYPQPNRPMGLNRWD  
DQRMETFMNSGVTVPVMPILTEHPPASPPGRGPESTERTTSRYEYSPSLEDID  
LIDVLRSDIAGEKGTQVAPADQYECDLQTLTEKSTVAPLTAEANARYEDLSKGFY  
GFESFNNOYQKHQOQREQIKTPLEHPTQAELEDDLFDAQLFEDVSRREG  
QLNQLFDNKKQHPVNNVLSSEGLVYNQANLTEMQMRDSCNOVISITPTSTAQPE  
TLFNVTDSQTEVQWLPEVVPNDVFTSNAYIGMONDSLQAVVNSGQIDYHSYQST  
GTPTPLIIGSGRQQTOTSPGVTVTATQSLFDPYHSRHSFDCCTTDSSTCS  
RLSSSPRYTSSESTGTHESRYGKLA PSSGRIQRSPSRSSQSIKTARVVPVPLAS  
GQRGRQKDEQLASDNLPLVSAFQISMSSELUQVLKNESEYQRLIRKRR  
GNKVAARTCRQRTDRHDKMSHYI"  
complement(join(36..112,165..532,686..944,1139..1367))  
/gene="skn-1"  
/note="coded for by the following C. elegans cDNAs:  
yk774f09.5, yk774f09.3"  
/codon\_start=1  
/product="C. elegans SKN-1 protein (corresponding sequence  
TL9E7.2b)"  
/protein\_id="AAM75365.1"  
/db\_xref="GI:21693909"  
/translation="MSLPSDEASSLLASSTTTTNTTAPAAVNSFDRQEEESKILNM  
YLQNMQQVQDGHGHHPYAYFGVSSFTDRVFTSNYAIQGNDSLQAVVNSGQI  
DDHSYQSTGTPTPLIIGSGRQQTOTSPGVTVTATQSLFDPYHSRHSFSD  
TATDSSSTCSRLSSSPRYTSSESTGTHESRYGKLA PSSGRIQRSPSRSSQSIK  
TARVVPVPLASGQRGRQKDEQLASDNLPLVSAFQISMSSELUQVLKNESEYQ  
RLIRKRRGNKVAARTCRQRTDRHDKMSHYI"  
2259..3224  
/gene="tl9e7.1"  
/note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=tl9e7.1;class=Sequence>"  
complement(join(2259..2414,2800..2950,2995..3224))  
/gene="tl9e7.1"  
/note="contains similarity to Pfam domain PF00059  
(lectin\_C), Score=82.4, E-value=2.9e-21, N=1; coded for by

CDS

gene

CDS



the following C. elegans cDNAs: CEESH64R, CEMSH13F"

```

/codon_start=1
/product="Hypothetical protein T19E7.1"
/protein_id="AA83595.1"
/db_xref="GI:125793"
/translation="MILLLLYLLIPTINVVIPSLKSSQSIAGORAINLDVYVRL
TVLALQSGWLYEDHYKMFETDYMWIPAEVNCVSMGHLVSIKDESENLFVHLK
KNIWIGLKLNDLPHVYKWDGSEADYLNWASSQNEPVDVDCAYNAFHQEGRTWED
YGCRLMPOFFVCKLMP"
/complement(9612..12689)
/gene="T19E7.3"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=T19E7.3;class
=Sequence"
/complement(11090..11637..11938..11985..12083..12591..12689))
/gene="T19E7.3"
/notes="coded for by the following C. elegans cDNAs:
YK391b8.5, YK486f7.5, YK610c10.5, YK656b6.5"
/codon_start=1
/product="Hypothetical protein T19E7.3"
/protein_id="AA83597.2"
/db_xref="GI:14574235"
/translation="MTQPSHICLNQHPRLDFTORRPSADSEKSEITVTEALTG
HSRLMKLISDAQFSDAPVNCDCSALRNEMDAQVTLDEIKTYQTVINYKLNHP
TTSIPDLKALQNVSDERELEQQLKLAEEQLDLDTQKRTAEASEKSEKWLK
KYRDLNRFVEQDELHSEAEQYAEVORHKLDTNVDLCFHWVGVGSENGFR
LGVLDKAPVETFEINAALGOVILLLELLERIGVQHVELPMVAMGSHSYIKLRNGID
METVLYGCTPLSGSGSDPGIRREFLLLEFLKELKDRNKNFKPPYQIHADSLVDN
GVKNVMTLNTDVRTRAMALMLTLKAAACQCDLRSPI"
/complement(20905..22799)

```

Alignment Scores:

Pred. No.:	427	Length:	23933
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	43.75%	Indels:	0
DB:	3	Gaps:	0

US-09-854-133-587 (1-16) x U42843 (1-23933)

QY 7 IleAspPheIleIlePheTyr 13  
 |||||  
 Db 9853 ATCGATTTATATTTTGG 9833

RESULT 41  
 U42841  
 LOCUS 31228 bp DNA linear INV 23-MAY-2002  
 DEFINITION Caenorhabditis elegans cosmid T17H7, complete sequence.  
 ACCESSION U42841  
 VERSION U42841.1 GI:2072756  
 KEYWORDS HTG.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.  
 1 (bases 1 to 31228)  
 Waterston, R.  
 Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium  
 Science 282 (5396), 2012-2018 (1998)  
 99069613  
 PUBMED 9851916  
 REFERENCE 2 (bases 1 to 31228)  
 Favello, A.  
 The sequence of C. elegans cosmid T17H7  
 JOURNAL Unpublished (2001)  
 3 (bases 1 to 31228)  
 REFERENCE  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JUL-2001) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

Louis, MO 63110, USA  
 4 (bases 1 to 31228)  
 Waterston, R.  
 Direct Submission  
 Submitted (14-SEP-2001) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
 Louis, MO 63110, USA  
 5 (bases 1 to 31228)  
 Waterston, R.  
 Direct Submission  
 Submitted (23-MAY-2002) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
 Louis, MO 63110, USA  
 On May 7, 1997 this sequence version replaced gi:1125779.  
 Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RQ, England  
 email: rw@nematoe.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we provide a small overlap between  
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
 regions were double stranded, sequenced with an alternate chemistry  
 or covered by high quality data (i.e., phred quality >= 30); an  
 attempt was made to resolve all sequencing problems, such as  
 compressions and repeats; all regions were covered by sequence from  
 more than one m13 subclone.

For a graphical representation of this cosmid sequence and its  
 analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=T17H7;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C24A1, 3897 bp overlap; the 3' cosmid is F42G9,  
 3000 bp overlap. Actual start of this cosmid is at base position  
 3698 of T17H7; actual end is at 19222 of F42G9.

NOTES:

Coding sequences below are the result of integration and manual  
 review of the following data: computer analysis using the program  
 Genefinder (P. Green and L. Hillier, personal communication), the  
 large scale EST projects of Yuji Kohara  
 ([http://www.dbbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.dbbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C.  
 elegans ORFeome cloning project (<http://worldb.dfci.harvard.edu/>),  
 similarity to other proteins from BlastX analyses  
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae  
 using Jim Kent's WABA alignment program (Genome Research  
 10:1115-1125, 2000), individual C. elegans GenBank submissions,  
 and personal communications with C. elegans researchers. tRNAs  
 are predicted using the program tRNAscan-SE (Lowe, T.M. and  
 Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES  
 source

1..31228  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /chromosome="III"  
 /clone="T17H7"  
 complement(52..18951)  
 /gene="T17H7.4"

gene

/note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=T17H7.4;class=s=Sequence>"  
 complement(join(52..118,361..477,530..839,1295..1457,  
 1506..1606,1993..2098,2399..2557,2722..2850,4319..4339)

CDS



```
mRNA
join(<400, .535,951, .1036,1480, .1590,2259, .2449,2909, .3061,
3729, .4273,4966, .5050,5682, .>5781)
/product="unknown"
CDS
join(<400, .535,951, .1036,1480, .1590,2259, .2449,2909, .3061,
3729, .4273,4966, .5050,5682, .5781)
/notes="CDS1"
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18870.1"
/db_xref="GI:20302743"
/translation="KAKRAQKGRCLPGVETVLEIIEQEEQRREKTRQAKRRESA
GSSQVLDPNTSFTTASEKDLSDVEGPTAESEETETPVPVLPKLELRE
HIMICRGLVLPQCPQNVITILENDRFAAGTSVSDRLHQLSRAAPFAKPEPT
PSNVHTRPVPPNVDLVKEVYGRIMDFILPLTLQYQDEEVQTMATSTFIPL
TPVAGTQDAATPPPPIKRVTRVSARSIPICHEAQKSHPEYPLRDSAVKFSQAL
SKSDGLSRRESFMDQEPALSLPTQKRRHSSISHAQPLPAARQAETQPSPPP
LIPAPPPEGPATADASKLSVDTAAEVKQVHAKLLDEAMCKLLPDFSODGON
PPLIYGAQLHLRLFKVLPPELLGKMLPEVYSKILCYHLQLFLKYLADHRDLLPATA
YVSAEVLGNLRQPKVIR"
complement(join(<6992, .7192,7693, .7832,8188, .8399,
8980, .9725,10309, .10443,10831, .10943,11221, .>11242))
/product="unknown"
complement(join(6992, .7192,7693, .7832,8188, .8399,
8980, .9725,10309, .10443,10831, .10943,11221, .11242))
/notes="CDS2"
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18871.1"
/db_xref="GI:20302744"
/translation="MAKFPVAGQADCPYAKVELLADLSQKLODFVHKIVKTPPEW
KPIVQLQCADNGWTHGSPILIRELVDRGGKGLIGHNDFMFAISYGMTSMLE
MKHIAAENLAVTOELREEEETIRQKPLKVCVINAARSAPYHVLPSLVNGKILREE
EIAHLHDSLENLEKLEMEVDFLSPFLKEISVTIDLETFQFQNAHIAVLDDFDQ
GKEDATGDMETKVSFYKRAEAINQTSKDIRVLVAGTGPLNSILVLIIDHTPSIPR
ONIAAQAQVKRQAKSLIAKRLTVNSAGVCDIVVNGVGGTYTIDVSRAVHGDGAI
WGPSYSCSVSEMYHDKWLEGELELOLSRSHITQDSLOHSDASMAAASSTLSYI
WNGSPGQLFLAVCSGEGFINPGEVSEFPYMHKGSWEVQDIDMNMVRLMSLI
TAQVLEEKDLHLHTNANHEKLLKVGFTIDTISFSSVKLKIEMEETKSESDTEPA
AAAAYSTGDETVPS"
join(<11840, .11941,12709, .12837,13348, .13456,14092, .14204,
14393, .14497,14775, .>14978)
/product="unknown"
join(11840, .11941,12709, .12837,13348, .13456,14092, .14204,
14393, .14497,14775, .14978)
/notes="CDS3"
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18872.1"
/db_xref="GI:20302745"
/translation="WAKSTSTWNTADVLPEKDKAKELRARIKIKQAEVLKPGAVF
VGHIPHGFEQVQKRRFQSGTVSVRLARSKTKGPGYGYIEFECEVAKVAETM
NLYLAKLKECFEPTERLHPKTFSGSNRKFLEKPTRRSTAIRKHNKPKDAATLOSA
MRLAKNLRKRKYAALGLNYDFSGYDDMAKAKLQAASIVTPKVPKSEEVKTKQE
PATLLSDSDPEISPKTPGVSVKIKKAAKQT"
join(<17025, .17173,17557, .17694,18464, .18964,
19413, .>19716)
/product="unknown"
join(17025, .17173,17557, .17694,18464, .18964,19413, .19716)
/notes="CDS4"
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18873.1"
/db_xref="GI:20302746"
/translation="MASETAPVWPQNTSPDKRKEGIRHRLVFLVLTLLCLSVIINA
LWLSRQGLTMLLARRDKNTDGNLRPVPPEVSEAPFSVEDVELRIRNAASKL
HDKRFLRKWPVSCLRKTKGKHKCPKPKIGNRRCAVVGSGILLGSDCGAD
IDAKDITRIDLPVPAFGEDYGRKTNMVLNMDTPKRLISSLRNKRLSDMTYHRLR
GTAGTVLLANRSQPYLKAQRIYHVSFVLLTSASSTSEKINPIASKIARVYMK
HPSTGMTAVLIATTFCDRLFLYGFPPGQDENKRPPIPHYIPDDGINQKFIADKMH
```

```
mRNA
MOTEFYFKRLKSSKGVKLKLHIRECDNO"
join(<20714, .20772,23503, .23669,24051, .24251,24572, .24624,
25001, .25018,25219, .25350,25808, .25850,26307, .26394,
26832, .26931,27374, .27604,27949, .28146,28746, .28940,
29352, .29543,30020, .30166,30774, .30848,31042, .31200,
32189, .32403,32972, .33200,33679, .>33891)
/product="unknown"
join(20714, .20772,23503, .23669,24051, .24251,24572, .24624,
25001, .25018,25219, .25350,25808, .25850,26307, .26394,
26832, .26931,27374, .27604,27949, .28146,28746, .28940,
29352, .29543,30020, .30166,30774, .30848,31042, .31200,
32189, .32403,32972, .33200,33679, .>33891)
/notes="CDS5"
/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAM18874.1"
/db_xref="GI:20302747"
/translation="MIRGLVCLLSLSSAVLAQQGPAYFISVPNLLKVGTEETVSNIF
NVPNPVRKVLYQDPDRKSTFSEAVDNDQDEPLVTVRVNADNLPSTATQRYVY
VAKSDDQLTFQKEAQVLLSQFGYVFTDKPIYTPNQKVMRIMPLDDQMTPASQ
VKLEILNPOGLIVFQKQFGSGTGFIAETDFPAPFLPGNMTTIAHYPENQLANSTQ
FEVREYVLPYGVKLIIPSNPYILPQDDVIGSEVALYTGKVGDFLKLGIIDQEG
NRQLFAQLQTEVNDGIGYIEIQTQRIKDLDLWFEPGSRSLYLEAAVTEAGGLREISVL
TSISFETSPFKIGYDLTATHEFKPGLFLVLTLPKPAEDIPVRSATAIIPGOD
PITVLGRNNEHNSDTNQGASFTVDVPPATQTLTVTAQTEQVGLPAAHQAEDEFA
TPYQSPASVLLVRQRPPIQDEAIDVAVTQKNDIQSYNVMVTRGVTHQGIK
VROGVLKTIITFTSVMNPISRLIVYINLQGEVAVDSTLLEIENVCNKVTVSSVD
NVEQEQANIEVADPPNSILGLLAVDQVAYLLNNYLTNLTOKMFOAMAKYDQCGCPG
GODSANVFKDAGVTVLTQVLSPAVRNAGCGTOSRRKRDQNSOEAQKSLFVCMVIT
VMIVRAKLNSTQDECYFAFLTCCHNARSRLQGRGRGMGGGLLIDIPIDDESOL
VARTEFETWTFEDVQDDGOAIVPVPISITTWIQAAGISTANGCMVAEPHME
SKFKFIHLQPLYSILQIEGVAVRATIFNYDQDLRVNVMYMGVSGVCSGARERSE
RKTFLIKGNDAAASVLPPIPLEVGFPIRVLAFTAAAGDITIEKSLQVI"
BASE COUNT 10151 a 7072 c 7010 g 10311 t
ORIGIN
Alignment Scores:
Pred. No.: 592 Length: 34544
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 3 Gaps: 0
US-09-854-133-587 (1-16) x AF391289 (1-34544)
QY 6 GlyIleaspPhellellePhe 12
|||||
Db 30200 GGGATGATTCATTATTTT 30220
RESULT 43
LOCUS L16621/c 37977 bp DNA linear INV 29-MAY-2002
DEFINITION Caenorhabditis elegans cosmid ZK688, complete sequence.
ACCESSION L16621
VERSION L16621.2 GI:15487650
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 37977)
Waterston,R.
Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
JOURNAL 98069613
MEDLINE 9851916
PUBMED 2 (bases 1 to 37977)
AUTHORS Wilson,R.
TITLE The sequence of C. elegans cosmid ZK688
JOURNAL Unpublished (2001)
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 37977)  
Waterston,R.  
Direct Submission  
Submitted (09-AUG-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
4 (bases 1 to 37977)  
Waterston,R.  
Direct Submission  
Submitted (07-SEP-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
5 (bases 1 to 37977)  
Waterston,R.  
Direct Submission  
Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
6 (bases 1 to 37977)  
Waterston,R.  
Direct Submission  
Submitted (29-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Sep 7, 2001 this sequence version replaced gi:289775.  
Submitted by:

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RO, England  
email: rwenematode.wustl.edu and jes@sanger.ac.uk

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate chemistry  
or covered by high quality data (i.e., phred quality >= 30); an  
attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by sequence from  
more than one ml3 subclone.

## COMMENT

For a graphical representation of this cosmid sequence and its  
analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=ZK688;class=Sequence>

## NEIGHBORING COSMID INFORMATION

The 5' cosmid is C02C2, 1600 bp overlap; the 3' cosmid is C39E4,  
1000 bp overlap. Actual start of this cosmid is at base position 1  
of ZK688; actual end is at 36977 of ZK688.

## NOTES:

Coding sequences below are the result of integration and manual  
review of the following data: computer analysis using the program  
GeneFinder (P. Green and L. Hillier, personal communication), the  
large scale EST projects of Yuji Kohara  
([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C.  
elegans ORFome cloning project (<http://worfdb.dcfi.harvard.edu/>),  
similarity to other proteins from BlastX analyses  
(<http://blast.wustl.edu/>), sequence conservation with C. briggsae  
(using Jim Kent's WABA alignment program (Genome Research  
10:1115-1125, 2000)), individual C. elegans GenBank submissions,  
and personal communications with C. elegans researchers. tRNAs  
are predicted using the program tRNAscan-SF (Lowe, T.M. and  
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).  
Location/Qualifiers  
1. 37977  
/organism="Caenorhabditis elegans"

FEATURES  
source

gene

/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="III"  
/clone="ZK688"  
complement(98..3813)  
/gene="ZK688.9"  
/note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=ZK688.9;class=Sequence>"  
complement(join(98..296,379..461,1044..1160,1294..1575,  
1684..1860,2714..2837,2886..3052,3097..3201,3778..3813))  
/gene="ZK688.9"  
/note="coded for by the following C. elegans cDNAs:  
yk657d4.3, yk657d4.5"  
/codon\_start=1  
/product="Hypothetical protein ZK688.9"  
/protein\_id="AAL00884.2"  
/db\_xref="GI:15967147"  
/translation="MSAKNRLDVKTRDVIKSLVDMKRGYKRNIGSRNQMKNV  
EKKVEOMGAKNEFMFSKIMKPKNSGHLQWSDQTLADLMNKSDSFDQESLIMP  
RSESSAMSVNRNTERQSEPVKSPRAAPKVNAGITATYVYSEMSDRSDEIVMV  
RAAKNAKREKFDLGFKISLAGHILESSCKDEQEDPSCKKCYDKRLBERPE  
MYFARNLSITQFRLGSIENFADALMVCADRLPDVKVGASTVQASDRDLQIOQISE  
HQPFDVTYTHYKGTVCQVPTTTERIDRLKRDEILESSITLFEDELADHGI  
AQLLARVYRMGYFFVLLRFYMRVDNVLRLVRCVTRIVGNEFDGHIKREKRYGN  
LGHVDPLELLDVRAMHLPVVEEHFDRYSVDRERLE"  
complement(4277..10419)  
/gene="ZK688.5"  
/note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=ZK688.5;class=Sequence>"  
complement(join(4277..4461,4657..4827,4889..5122,  
5283..5975,6021..6213,6638..8566,8618..9006,9053..9315,  
9416..9622,9674..9985,10085..10251,10300..10419))  
/gene="ZK688.5"  
/note="contains similarity to Pfam domain PF00240  
(ubiquitin), Score=14.4, E-value=0.012, N=1; coded for by  
the following C. elegans cDNAs: yk16d10.3, yk16d10.5,  
yk36d7.3, yk36d7.5, yk39d2.3, yk39d2.5, yk15b5.5,  
yk161a3.5, yk171c4.5, yk171c4.3, yk493c6.5, yk484g03.3"  
/codon\_start=1  
/product="Hypothetical protein ZK688.5"  
/protein\_id="AAL00883.1"  
/db\_xref="GI:15487652"

gene

/translation="MSETQEQEASNGCEPDLPTTIRVTLKTLDDREAVTIGLQDTIQ  
SLDLGRRENNIQSGFORVIAGRVLSSTQVQAIGSDGQTVHLVDGSPGSENDPRN  
VMPDRVAGPRIINAIPLGPPGFIFQSPAFARMIPGNVEIPTPSPQTHVVRVIRVP  
GSIAGTPVLTSRYSEDCVLQKAVPIRGTSNPNPQAASQVTFPSENPVQMTVYIA  
DLILFRPREHEQVVRTINNISLSDSTRLGVSMKNQNTSLSVLPSPHPIPS  
ALEKDLFLCWLTDHLSDIKLEEHDGLVAATRHVLEMKTRQFDONLSQOARNORVE  
ADDEIVKHLEYOWAELSHMKDFERIRFKQTEYERALKIQEYETPNPHFYMRHAL  
DIDVIGMREFRQQRFTLEDLLDNEVGAVKEIRDTTTEVDYRQALSMFYCY  
IQMRHQISHTHLTADLVSFITPSPQRLIQYAFQSSDEQELTIPVTHIFPPRM  
LSDVGGYNGDYPYLAHPBSVMEVLOQPRIAEPRLAQTTPQQLVLTQMNMQ  
GGSINLMAATDPLISGLQDLVLAQEQVENLFAQOPIEGGRVVRVARGPGRVATT  
GDAPSVPTLPPGSDQQRRTTHRFNVQPDAGAEETLAPFPAIEFNELO  
RIAKNIASRIAEALQIASLITFRNNSWDTRIQNMPLCTLRCAVIALEMTLSAC  
NTVNSKDLMLALVDEEVLVRAIAEIKSLFGKGEPTTHVILMTFVLEMLTSG  
ESVDGVEESQSDFSMIVRPSDISQPGSDILRAIRERQNRQFLNENGRIFSTSS  
APSTENPGPSFNSDDAIDIRAGLPIDGTIRNRVREIHFVFAAARASPNHISIT  
FTATHTPAPAGFPPLMMASSNPSTAGPGWFIQVVSPTTHGFLFELDLSSGSDQ  
PARSPPTAPPTPRTSAPATVQSSPTRESMDISPNQVONPGHVESPAIAAQAAR  
VARADILHANTFNGDLARSQSPFVPTPTPLNDPRRTVRVTRVQHVAPMAIDPF  
MCTNRHCEINLPTFTHMADGRTFLOLPDQFEARQALVSTERRPIOHIED  
EDYNSIRRTQSLGSLFRNLEDPRFVKTAIRSLAHCIDADITFLFMNMMNISGYQAA  
NHTLELMVKHISRVPGSARTRNASONTTSVNOSTETSPREOMKNSPVQEAADPRIA  
FSPPTPLNQEQEVLIPQIASLITLYVMESSNPFRPPGIFGLLELTGLRLTRHD  
FAOLARRTTATNVAESEFAQIRAHIRDNLTGRTGLNSSELHGIAENLNEOFFAIF  
MSQNDQLTSPFGVDRNDFAEVWAFQIETALIKSFTLSQNLKDSNNAVFILQOS  
VDSILRNLMIFRMCDRDVERMKIQRISYFATIRYESTDRPGINVFIDNNRMV  
DWSNRYTDFSEENFDQFLKVRAGTDWNDIVLNSRQLLPTIASSSPSPSSSSSLIN  
TVSTNNPRKRNHDSREDNDDCVVAVPMTLSLSSPLTSQSSSSSGTITSSGSGPS  
TSSSTTTNNIQ"

CDS

```

gene
12630..13366
/contig="ZK688.3"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=ZK688.3;class
=Sequence"
join(12630..12752,12797..12904,12953..13366)
/contig="ZK688.3"
/note="contains similarity to Pfam family PF01557
(Fumarylacetylacetate (FAA) hydrolase family), score=340.3,
E=2.1e-98, N=1; coded for by the following C. elegans
cDNAs: yk899e08.5, CEESP53F, CEESP6F, yk899e08.3"
/codon_start=1
/product="Hypothetical protein ZK688.3"
/protein_id="AAA28230.1"
/db_xref="GI:289782"
/translation="MSSLAGFRNLATKIVCVGRNFKDALELGNAPKPKMLFVKTVN
SFVEGPIVAPGCONLHVEVLGVVSKKASRISKSDAMDYIGGYTVALDMDARDF
QDEAKAGAPFLAKSFDGCPGIGFLPVSDLPNPHDVELFKNGDQQRCTDVM
FDIPLLEYTQFFTFLEVDVVLGTGTPAGVTKINSGDVIEFLGLDGLNSKFNQ"
13883..14129
/contig="ZK688.4"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=ZK688.4;class
=Sequence"
join(13883..14027,14071..14129)
/contig="ZK688.4"
/codon_start=1
/product="Hypothetical protein ZK688.4"
/protein_id="AAA28229.1"
/db_xref="GI:289781"
/translation="MHTLLFRTVHSVNFYIFSTILLTSAVILAIFFFFGDGFRE
KYRFEFTNSKEKGNKNTIVE"
19891..19892
/misc_feature
/note="S11 trans-splice acceptor; see yk729b7.5"

Alignment Scores:
Pred. No.: 644 Length: 37977
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x L16621 (1-37977)
QY 6 GlycylAspPhelIlePhe 12
|||||
Db 17672 GGGATTGACTTATCATTTT 17652

RESULT 44
AC002499 41052 bp DNA linear PRI 10-OCT-1997
LOCUS Human Cosmid g5129s435 from 7q31.3, complete sequence.
DEFINITION AC002499
ACCESSION AC002499
VERSION AC002499.1 GI:2342698
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 41052)
Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,
Green,P. and Olson,M.V.
Large-scale MCD Mapping and Sequencing of Human Chromosome 7
Unpublished (1996)
JOURNAL University of Washington Human Genome Center
REMARK Box 352145 Seattle, WA 98195
Contact: Shawn Iadonato (iadonato@u.washington.edu)
2 (bases 1 to 41052)
Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,
Green,P. and Olson,M.V.
Direct Submission
Submitted (25-AUG-1997) Human Genome Center, University of

```

```

REMARK Washington, Box 352145, Seattle, WA 98195, USA
University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Shawn Iadonato (iadonato@u.washington.edu)
REFERENCE
3 (bases 1 to 41052)
Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,
Green,P. and Olson,M.V.
Direct Submission
Submitted (10-OCT-1997) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
JOURNAL University of Washington Human Genome Center
REMARK Box 352145 Seattle, WA 98195
Contact: Shawn Iadonato (iadonato@u.washington.edu)
COMMENT Overlapping Sequences:
5': UWGC:g5129g101
3': UWGC:g5129g124
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Pct of Consensus above quality 40: 97.9%
Number of manually edited bases: 7
Double stranded (DS) coverage: 91.0%
DS or two chemistry coverage: 7.0%
Single stranded regions: 3
Quality Cumulative Percentage of Consensus
-----
90 xx ( 8.3%)
80+ xxxxxxxx (40.6%)
70+ xxxxxxxxxxxxxxxx (76.2%)
60+ xxxxxxxxxxxxxxxxxxxx (88.3%)
50+ xxxxxxxxxxxxxxxxxxxx (94.3%)
40+ xxxxxxxxxxxxxxxxxxxx (97.9%)
30+ xxxxxxxxxxxxxxxxxxxx (99.6%)
20+ xxxxxxxxxxxxxxxxxxxx (100.0%)
10+ xxxxxxxxxxxxxxxxxxxx (100.0%)
00+ xxxxxxxxxxxxxxxxxxxx (100.0%)
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Validation:
This sequence has been validated by Multiple Complete Digest
Mapping. Comparison of the experimentally derived map digest
fragments with sequence-predicted fragments is given below.
Small fragments below a variable cutoff (approximately 400-600bp)
are not mapped and hence do not appear in the table. There are no
significant remaining discrepancies between the experimental and
predicted values. Uniquely ordered fragment groups are separated
by dashed lines.
EcoRI HindIII
Map Seq Map Seq Map Seq
-----
3714.42 3709.00 1212.81 1206.00 11295.40 11398.00
-----
5425.43 5420.00 1591.62 1586.00 2802.45 2811.00
-----
13174.96 13413.00 1902.56 1912.00 16219.29 16613.00
-----
8209.13 8223.00 10264.00 10325.00 3557.24 3563.00
-----
3745.48 3711.00 5799.89 5846.00
-----
4214.65 4237.00
-----
1794.23 1814.00
-----

```

```

-----
2948.97 2950.00
-----
3981.14 3989.00
-----
2571.06 2548.00
-----
1350.02 1384.00
-----
1349.38 1345.00
-----
2396.63 2389.00
-----
-----
FEATURES
source
Location/Qualifiers
1. .41052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31.3"
/clone="NHGRI:YWSS5129 or CEPH:744C02"
/sub_clone="UMGC:g5129s435"
/clone_lib="CEPH MegayACs"
262. .621
/rpt_family="THE1"
complement(2652. .2887)
/rpt_family="ALU"
3512. .3773
/rpt_family="ALU"
complement(4757. .7115)
/rpt_family="L1"
complement(6137. .6430)
/rpt_family="ALU"
complement(7130. .7441)
/rpt_family="ALU"
complement(7481. .7832)
/rpt_family="L1"
10859. .11147
/rpt_family="ALU"
12294. .12704
/rpt_family="MER25"
12816. .13474
/rpt_family="L1"
13543. .13629
/rpt_family="L1"
14610. .14841
/rpt_family="ALU"
16511. .16794
/rpt_family="ALU"
17402. .17693
complement(20329. .21547)
/rpt_family="L1"
23252. .23361
/rpt_family="ALU"
24813. .27565
/rpt_family="L1"
27809. .28105
/rpt_family="ALU"
29332. .29518
/rpt_family="ALU"
complement(30500. .30604)
/rpt_family="ALU"
complement(31397. .31779)
/rpt_family="THE1"
complement(32677. .33025)
/rpt_family="THE1"
complement(33043. .34521)
/rpt_family="MSTAR"
complement(34545. .34936)
/rpt_family="THE1"
complement(35281. .35491)
/rpt_family="MTR"
complement(35935. .37518)

```

```

repeat_region /rpt_family="L1"
complement(37523. .39574)
/rpt_family="L1"
BASE COUNT 13425 a 7256 c 7164 g 13207 t
ORIGIN
Alignment Scores:
Pred. No.: 691 Length: 41052
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 9 Gaps: 0
US-09-854-133-587 (1-16) x AC002499 (1-41052)
QY 9 PheIIeIIePheTrIpIlePhe 15
Db 40866 TTTATCATATTTGGATATTT 40886
RESULT 45
AC003032
LOCUS AC003032 42354 bp DNA linear PRI 28-OCT-1997
DEFINITION Human Chromosome Y Cosmid 7A1 Genomic Sequence, complete sequence.
ACCESSION AC003032
VERSION AC003032.1 GI:2570226
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 42354)
AUTHORS Thorstenson, Y., Au, M., Chung, E., Hyman, R.W., Schramm, S.,
Vollrath, D. and Davis, R.W.
TITLE Human Chromosome Y Cosmid 7A1 Genomic Sequence
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 42354)
AUTHORS Thorstenson, Y., Au, M., Chung, E., Hyman, R.W., Schramm, S.,
Vollrath, D. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1997) Biochemistry, Stanford University/DNA
Sequencing and Technology Center, 855 California Avenue, Palo Alto,
CA 94304, USA
COMMENT For each base, there are (at least) three quality reads, with at
least one read from each strand. A serious attempt at
editing/finishing has been made to resolve all sequencing problems,
such as guanine compressions.
Cosmid 7A1 is part of a three cosmid contig, with cosmids 54E8 and
56B5.
FEATURES
Location/Qualifiers
1. .42354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/cell_line="human/hamster J640-51"
/clone_lib="L10NCO3 'M'"
/note="The library was constructed by J. Garnea and P. de
Jong at Lawrence Livermore National Laboratory, using
flow-sorted Y chromosomes. Vector: Laurist 16."
BASE COUNT 13578 a 8319 c 8145 g 12312 t
ORIGIN
Alignment Scores:
Pred. No.: 710 Length: 42354
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 9 Gaps: 0
US-09-854-133-587 (1-16) x AC003032 (1-42354)
QY 9 PheIIeIIePheTrIpIlePhe 15

```

us-09-854-133-587.oligo.rge

• Tue May 13 12:12:57 2003

Db 31661 TTTATTTCTCGGATTTT 31681  
|||||

Search completed: May 11, 2003, 17:20:51  
Job time : 424.071 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 16:14:30 ; Search time 34.2655 Seconds  
(without alignments)  
1051.553 Million cell updates/sec

Title: US-09-854-133-587

Perfect score: 16  
Sequence: 1 PQANGIDFIIFWIFW 16

Scoring table:

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4348058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters:

-MODEL=framer\_p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO\_spool/US09854133/runat\_05052003\_174132\_594/app\_query.fasta\_1.462  
-DB=N\_Geneseq\_101002 -OPMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=60 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=45 -MODE=LOCAL  
-OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn2.1.275 @runat\_05052003\_174132\_594 -ICPU=3  
-NO\_XLPHY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -XGAPOP=60 -XGAPEXT=6 -FGAPEXT=7  
-WARN\_TIMEOUT=30 -THREADS=1 -DELOP=60 -DELOP=6 -DELEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	337	23	Human lung tumour-
2	16	100.0	2239	23	Human lung tumour-
3	16	100.0	5981	23	Human lung tumour-
c 4	7	43.8	579	24	Human cancer relat
c 5	7	43.8	826	16	Flower style-speci
c 6	7	43.8	826	18	Tomato S-ribonucle
c 7	7	43.8	2241	18	Tomato S-ribonucle
c 8	7	43.8	5886	24	Human immune syste
9	7	43.8	8576	24	Human immune syste
10	7	43.8	9265	24	Human immune syste
11	7	43.8	9265	24	Chemically pretrea
12	7	43.8	15723	23	Drosophila melanog
c 13	7	43.8	15881	24	Chemically treated
14	7	43.8	15881	24	Human immune syste
15	7	43.8	15881	24	Human gene regulat
16	7	43.8	15881	24	Signal transductio
17	7	43.8	16633	24	Human chemically m
18	6	37.5	26	19	3' PCR primer used
c 19	6	37.5	246	21	Human secreted pro
c 20	6	37.5	250	24	Human ovarian canc
21	6	37.5	277	22	Human breast cance
22	6	37.5	277	22	Human breast cance
23	6	37.5	277	22	Human breast cance
24	6	37.5	283	24	Human ovarian canc
c 25	6	37.5	303	22	Human immune/haema
c 26	6	37.5	312	23	Human prostate exp
c 27	6	37.5	339	23	Human prostate exp
c 28	6	37.5	416	23	Human prostate exp
29	6	37.5	418	20	DNA encoding novel
30	6	37.5	498	24	EST clone DK70. H
31	6	37.5	530	24	Gene #1080 used to
c 32	6	37.5	530	24	Oligonucleotide fo
c 33	6	37.5	530	24	Oligonucleotide fo
34	6	37.5	548	22	Human cDNA clone (
35	6	37.5	553	24	Oligonucleotide fo
36	6	37.5	553	24	Oligonucleotide fo
c 37	6	37.5	559	24	Oligonucleotide fo
c 38	6	37.5	572	22	Oligonucleotide fo
39	6	37.5	572	22	Human secreted pro
40	6	37.5	579	22	Human musculoskele
41	6	37.5	582	23	DNA encoding novel
c 42	6	37.5	584	21	Human secreted pro
43	6	37.5	600	22	Human foetal liver
44	6	37.5	600	22	Probe #8345 for ge
45	6	37.5	600	22	Human brain expres
46	6	37.5	600	22	Human bone marrow
47	6	37.5	600	22	Probe #11222 used
48	6	37.5	600	22	Human genome-deriv
49	6	37.5	600	24	Human colon cancer
50	6	37.5	675	24	GIP6-22 coding seq
c 51	6	37.5	696	22	Bacillus clausii g
c 52	6	37.5	696	24	Human cervical can
c 53	6	37.5	740	22	Human neuroblastom
54	6	37.5	748	22	Oligonucleotide fo
55	6	37.5	764	24	Oligonucleotide fo
c 56	6	37.5	801	22	Clostridium tetani
c 57	6	37.5	802	22	Human cDNA clone (
c 58	6	37.5	822	21	Plasmodium falcipa
c 59	6	37.5	831	24	Oligonucleotide fo
c 60	6	37.5	831	24	Oligonucleotide fo

## ALIGNMENTS

RESULT 1  
AAD23462  
ID AAD23462 standard; cDNA; 337 BP.

XX AC AAD23462;  
 XX DT 26-FEB-2002 (first entry)  
 XX DE Human lung tumour-specific 20E10 5' cDNA.  
 XX KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 KW antisense-therapy; vaccine; immune response; lung cancer; 20E10; ss.  
 XX OS Homo sapiens.  
 XX PN WO200172295-A2.  
 XX PD 04-OCT-2001.  
 XX PF 28-MAR-2001; 2001WO-US09991.  
 XX PR 29-MAR-2000; 2000US-0538037.  
 XX PR 05-JUN-2000; 2000US-0588937.  
 XX PR 18-AUG-2000; 2000US-0640878.  
 XX PR 22-SEP-2000; 2000US-234517P.  
 XX PR 01-NOV-2000; 2000US-0704512.  
 XX PR 14-DEC-2000; 2000US-0738973.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX WPI; 2001-639201/73.  
 XX DT New human lung-specific polynucleotides and polypeptides for the  
 XX diagnosis and treatment of disease e.g. lung cancer -  
 PS Claim 1; Page 334; 378pp; English.  
 XX CC The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer. The  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is a cDNA encoding human lung tumour-specific protein.  
 XX SQ Sequence 337 BP; 103 A; 60 C; 93 G; 81 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.79e-09 Length: 337  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0  
 US-09-854-133-587 (1-16) x AAD23462 (1-337)  
 QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
 Db 107 TTCCAGGCCAATTGTGGCATAGATTTTATCATATTCGATTTTGG 154  
 RESULT 2  
 AAD23460  
 ID AAD23460 standard; cDNA; 2239 BP.  
 XX AC AAD23460;  
 XX DT 26-FEB-2002 (first entry)  
 XX DE Human lung tumour-specific 19A4 cDNA.

KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 KW antisense-therapy; vaccine; immune response; lung cancer; 19A4; ss.  
 XX OS Homo sapiens.  
 XX PN WO200172295-A2.  
 XX PD 04-OCT-2001.  
 XX PF 28-MAR-2001; 2001WO-US09991.  
 XX PR 29-MAR-2000; 2000US-0538037.  
 XX PR 05-JUN-2000; 2000US-0588937.  
 XX PR 18-AUG-2000; 2000US-0640878.  
 XX PR 22-SEP-2000; 2000US-234517P.  
 XX PR 01-NOV-2000; 2000US-0704512.  
 XX PR 14-DEC-2000; 2000US-0738973.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX WPI; 2001-639201/73.  
 XX DT New human lung-specific polynucleotides and polypeptides for the  
 XX diagnosis and treatment of disease e.g. lung cancer -  
 PS Claim 1; Page 332; 378pp; English.  
 XX CC The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer. The  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is a cDNA encoding human lung tumour-specific protein.  
 XX SQ Sequence 2239 BP; 619 A; 444 C; 493 G; 683 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 9.06e-09 Length: 2239  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0  
 US-09-854-133-587 (1-16) x AAD23460 (1-2239)  
 QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
 Db 104 TTCCAGGCCAATTGTGGCATAGATTTTATCATATTCGATTTTGG 151  
 RESULT 3  
 AAD23461  
 ID AAD23461 standard; cDNA; 5981 BP.  
 XX AC AAD23461;  
 XX DT 26-FEB-2002 (first entry)  
 XX DE Human lung tumour-specific 14F10 full length cDNA.  
 XX KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 KW antisense-therapy; vaccine; immune response; lung cancer; 14F10; ss.  
 XX OS Homo sapiens.  
 XX PN WO200172295-A2.  
 XX

```
PD 04-OCT-2001.
XX
PF 28-MAR-2001; 2001WO-US09991.
XX
PR 29-MAR-2000; 2000US-0538037.
XX
PR 05-JUN-2000; 2000US-0588937.
XX
PR 18-AUG-2000; 2000US-0640878.
XX
PR 22-SEP-2000; 2000US-234517P.
XX
PR 01-NOV-2000; 2000US-0704512.
XX
PR 14-DEC-2000; 2000US-0738973.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX
DR WPI; 2001-639201/73.
XX
XX New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
XX
XX Claim 1; Page 332-334; 378pp; English.
XX
CC The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a cDNA encoding human lung tumour-specific protein.
XX
SQ Sequence 5981 BP; 1858 A; 1029 C; 1098 G; 1996 T; 0 other;

Alignment Scores:
Pred. No.: 2,1e-08 Length: 5981
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x AAD23461 (1-5981)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
DB 102 TTCAGGCCAATTGGCATAGATTTTATCATATCTCGATTTTGG 149

RESULT 4
ABN65138/c
ID ABN65138 standard; cDNA; 579 BP.
XX
AC ABN65138;
XX
XX 28-JUN-2002 (first entry)
XX
DE Human cancer related polynucleotide SEQ ID NO 5105.
XX
KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW gene therapy; cancer; tumour; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200214500-A2.
XX
XX 21-FEB-2002.
XX
XX 16-AUG-2001; 2001WO-US25840.
XX
XX 16-AUG-2000; 2000US-226326P.
XX
XX (CHIR) CHIRON CORP.
PA

(PA) (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX
XX WPI; 2002-241905/29.
XX
XX New nucleic acid for producing a polypeptide, detecting differentially
PT expressed genes correlated with a cancerous state of a mammalian cell,
PT and inhibiting tumor growth -
XX
XX Claim 1; SEQ ID NO 5105; 883pp + Sequence Listing; English.
XX
XX The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC with cytostatic activity. The polynucleotide is used to produce a
CC polypeptide, to detect differentially expressed genes correlated with a
CC cancerous state of a mammalian cell and to inhibit tumour growth. The
CC polynucleotide is used as a probe in mapping and tissue profiling. The
CC encoded polypeptide and antibodies to the polypeptide can also be used
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC gene therapy.
XX
XX Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 579 BP; 195 A; 123 C; 100 G; 161 T; 0 other;

Alignment Scores:
Pred. No.: 22.5 Length: 579
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABN65138 (1-579)

QY 10 IleIlePheTrpIlePheTrp 16
DB 328 ATAAATATTTGGATATTTGG 308

RESULT 5
AAQ79736/c
ID AAQ79736 standard; cDNA; 826 BP.
XX
AC AAQ79736;
XX
XX 21-SEP-1995 (first entry)
XX
DE Flower style-specific tomato S-ribonuclease.
XX
KW S-ribonuclease; flower style-specific; self-incompatibility; ds.
XX
OS Lycopersicon peruvianum.
XX
XX Key Location/Qualifiers
XX CDS 1..489
XX FT /*tag= a
XX FT /note= "START codon absent"
XX
XX JP06335389-A.
XX
XX 06-DEC-1994.
XX
XX 27-MAY-1993; 93JP-0126286.
XX
XX 27-MAY-1993; 93JP-0126286.
XX
XX (KAGO) KAGOME KK.
XX
XX WPI; 1995-056901/08.
XX
XX P-PSDB; AAR66604.
XX
```

PT Flower style-specific S-ribonuclease and DNA coding sequence -  
 XX isolated from self-incompatible wild tomato species  
 PS Claim 4; Pages 5-6; 10pp; Japanese.  
 XX  
 CC AAQ79736 encodes AAR66604 a flower style-specific S-ribonuclease,  
 CC isolated from a self-incompatible wild tomato species.  
 XX  
 SQ Sequence 826 BP; 309 A; 150 C; 131 G; 236 T; 0 other;

Alignment Scores:  
 Pred. No.: 30.5 Length: 826  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 100.00%  
 DB: 43.75%  
 Indels: 0  
 Gaps: 0

US-09-854-133-587 (1-16) x AAQ79736 (1-826)

QY 6 GlyIleAspPheIleIlePhe 12  
 Db 550 GGAATTGATTTATTATATT 530

## RESULT 6

AAT6453/c  
 ID AAT64553 standard; cDNA to mRNA; 826 BP.  
 XX  
 AC AAT64553;

DT 22-MAY-1997 (first entry)

DE Tomato S-ribonuclease cDNA fragment.

KW S-ribonuclease; plant style; promoter; tissue-specific;  
 KW self-incompatibility; ds.

XX Lycopersicon peruvianum.

XX Key Location/Qualifiers  
 FT CDS 1..489  
 FT /\*tag= a

XX JP09028381-A.

XX 04-FEB-1997.

XX 24-JUL-1995; 95JP-0187557.

XX 24-JUL-1995; 95JP-0187557.

XX (CHEN/) CHENG E K.

XX (MINA/) MINAMI H.

XX (TAKA/) TAKAGI M.

XX WPI; 1997-159091/15.

XX P-PSDB; AAW14912.

XX Tomato S-ribonuclease gene and promoter sequence - useful for  
 PT expressing genes in plant style cells, e.g. to confer  
 PT self-incompatibility on the plant

XX Claim 4; Page 10-11; 15pp; Japanese.

XX The present sequence encodes part of a tomato (Lycopersicon  
 CC peruvianum) S-ribonuclease. The S-gene can confer self-  
 CC incompatibility on a plant.

XX Sequence 826 BP; 309 A; 150 C; 131 G; 236 T; 0 other;

Alignment Scores:  
 Pred. No.: 30.5 Length: 826  
 Score: 7.00 Matches: 7

Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 43.75%  
 DB: 18  
 Gaps: 0

US-09-854-133-587 (1-16) x AAT64553 (1-826)

QY 6 GlyIleAspPheIleIlePhe 12  
 Db 550 GGAATTGATTTATTATATT 530

## RESULT 7

AAT64552/c

ID AAT64552 standard; DNA; 2241 BP.

XX AAT64552;

XX 22-MAY-1997 (first entry)

XX Tomato S-ribonuclease gene S12 allele.  
 DE  
 XX S-ribonuclease; plant style; promoter; tissue-specific;  
 KW self-incompatibility; ds.

XX Lycopersicon peruvianum.

XX Key Location/Qualifiers

FT promoter 1..1365

FT /\*tag= a

FT /note= "plant style-specific"

FT TATA\_signal 1255..1258

FT CDS 1366..1596

FT /\*tag= c

FT intron 1597..1687

FT /\*tag= d

FT CDS 1688..2101

FT /\*tag= e

XX JP09028381-A.

XX 04-FEB-1997.

XX 24-JUL-1995; 95JP-0187557.

XX 24-JUL-1995; 95JP-0187557.

XX (CHEN/) CHENG E K.

XX (MINA/) MINAMI H.

XX (TAKA/) TAKAGI M.

XX WPI; 1997-159091/15.

XX P-PSDB; AAW14911.

XX Tomato S-ribonuclease gene and promoter sequence - useful for  
 PT expressing genes in plant style cells, e.g. to confer  
 PT self-incompatibility on the plant

XX Claim 2; Page 9-10; 15pp; Japanese.

XX The present sequence is the tomato S12 gene from Lycopersicon  
 CC peruvianum coding for an S-ribonuclease; the S12 promoter region  
 CC has activity in plant style tissue. The S-gene can confer self-  
 CC incompatibility on a plant. In addition, a desired DNA sequence  
 CC can be expressed specifically in style cells of a plant using the  
 CC S12 promoter sequence.

XX Sequence 2241 BP; 583 A; 390 C; 444 G; 724 T; 0 other;

Alignment Scores:  
 Pred. No.: 71.7 Length: 2241  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00%  
 Gaps: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 18 Gaps: 0

US-09-854-133-587 (1-16) x ANT64552 (1-2241)

QY 6 GlyileAspPheIleIlePhe 12  
 ID ABL34212  
 XX  
 AC ABL34212;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 2183

RESULT 8  
 ABL34212  
 ID ABL34212 standard; DNA; 5886 BP.  
 XX  
 AC ABL34212;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 2185.

Human; immune system disease; cytosine methylation; antiasthmatic;  
 antiarteriosclerotic; antianemic; cytostatic; nootropic;  
 neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

PS WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful  
 for diagnosis and treatment of diseases associated with abnormal  
 cytosine methylation

PS Claim 1; SEQ ID NO 2185; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated  
 genes which are modified by the methylation of cytosines. The sequences  
 can be used in the diagnosis and treatment of immune system disorders,  
 including eye diseases such as retinopathy, neovascular glaucoma and  
 macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 diseases. The present sequence is a gene of the invention.

XX Sequence 5886 BP; 1482 A; 47 C; 1292 G; 3065 T; 0 other;

Alignment Scores:  
 Pred. No.: 164 Length: 5886  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABL34212 (1-5886)

QY 6 GlyileAspPheIleIlePhe 12  
 ID ABL34229  
 XX  
 AC ABL34229;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 2202.

RESULT 9  
 ABL34229  
 ID ABL34229 standard; DNA; 8576 BP.  
 XX  
 AC ABL34229;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 2202.

Human; immune system disease; cytosine methylation; antiasthmatic;  
 antiarteriosclerotic; antianemic; cytostatic; nootropic;  
 neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

PS WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful  
 for diagnosis and treatment of diseases associated with abnormal  
 cytosine methylation

PS Claim 1; SEQ ID NO 2202; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated  
 genes which are modified by the methylation of cytosines. The sequences  
 can be used in the diagnosis and treatment of immune system disorders,  
 including eye diseases such as retinopathy, neovascular glaucoma and  
 macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 diseases. The present sequence is a gene of the invention.

XX Sequence 8576 BP; 2122 A; 64 C; 1723 G; 4667 T; 0 other;

Alignment Scores:  
 Pred. No.: 226 Length: 8576  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABL34229 (1-8576)

QY 6 GlyileAspPheIleIlePhe 12  
 ID ABL34153  
 XX  
 AC ABL34153;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 2203

RESULT 10

ABL34153

ID ABL34153 standard; DNA; 9265 BP.  
 AC ABL34153;  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX Human immune system associated gene SEQ ID NO: 2126.  
 DE  
 XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytosinetic; cytosinetic; cytosinetic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriasis;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200200928-A2.  
 PN  
 XX 03-JAN-2002.  
 PD  
 XX 02-JUL-2001; 2001WO-EP07537.  
 PF  
 XX 30-JUN-2000; 2000DE-1032529.  
 PR  
 XX 01-SEP-2000; 2000DE-1043826.  
 PR  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 XX WPI; 2002-130909/17.  
 DR  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 XX Claim 1; SEQ ID NO 2126; 32pp + Sequence Listing; German.  
 PS  
 XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 XX Sequence 9265 BP; 2686 A; 133 C; 1801 G; 4645 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 242 Length: 9265  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-854-133-587 (1-16) x ABL34153 (1-9265)  
 OY 9 PheillelePheTrpIlePhe 15  
 Db 525 TTTATTTATTTTGGATATTT 545  
 RESULT 11  
 AAS63355  
 ID AAS63355 standard; DNA; 9265 BP.  
 XX  
 AC AAS63355;  
 XX  
 XX 29-JAN-2002 (first entry)  
 DT  
 XX

DE Chemically pretreated metabolism associated gene #50.  
 XX  
 KW Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;  
 KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;  
 KW single nucleotide polymorphism detection; SNP; stool; urine; lung;  
 KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast;  
 KW DUSP2; EPHX2; QDPR; SSSH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200176451-A2.  
 PN  
 XX 18-OCT-2001.  
 PD  
 XX 06-APR-2001; 2001WO-EP04016.  
 PF  
 XX 06-APR-2000; 2000DE-1019058.  
 PR  
 XX 07-APR-2000; 2000DE-1019173.  
 PR  
 XX 30-JUN-2000; 2000DE-1032529.  
 PR  
 XX 01-SEP-2000; 2000DE-1043826.  
 PR  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 XX WPI; 2002-010834/01.  
 DR  
 XX New nucleic acid, useful for diagnosis and therapy of metabolic  
 PT disease, solid tumour and cancers, comprises segment of chemically  
 PT modified genomic sequences of genes associated with metabolism -  
 XX  
 XX Claim 1; Page 129-132; 143pp; English.  
 PS  
 XX The invention relates to a nucleic acid (I) comprising a sequence at  
 CC least 18 bases of a segment of the chemically pretreated DNA of genes  
 CC associated with metabolism such as DUSP2 (NM\_004418), EPHX2 (NM\_001979),  
 CC QDPR (NM\_000320), SSSH (NM\_000199), SHMT2 (NM\_005412), SLC7A2  
 CC (NM\_003046), SLC7A4 (NM\_004173) and TYMS (NM\_001071) (all  
 CC undefined). (I) are useful for diagnosis and therapy of metabolic  
 CC disease, solid tumours and cancers; as primer oligonucleotides for the  
 CC amplification of DNA sequences, for detecting the cytosine methylation  
 CC state and/or single nucleotide polymorphisms (SNPs) in a chemically  
 CC treated DNA of genes associated with metabolism. An array of (I) is  
 CC useful for ascertaining genetic and/or epigenetic parameters for the  
 CC diagnosis and/or therapy of existing diseases or the predisposition to  
 CC specific diseases by analysing cytosine methylations. The method involves  
 CC chemically treating genomic DNA sample by a solution of bisulphite,  
 CC hydrogen sulphite or disulphite such that cytosine bases which are  
 CC unmethylated at the 5th-position are converted to uracil or another base  
 CC which is dissimilar to cytosine in terms of hybridisation behaviour and  
 CC amplifying fragments of the chemically pretreated genomic DNA. The  
 CC genomic DNA is from cells or cellular components which contain DNA,  
 CC sources of DNA comprising, for e.g. cell lines, biopsies, blood, sputum,  
 CC stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as  
 CC tissue from eye, intestine, kidney, brain, heart, prostate, lung, breast  
 CC or liver, histologic object slides and their combinations. Genetic  
 CC parameters are mutations, in particular insertions, deletions, point  
 CC mutations, inversions and polymorphisms of genes associated with  
 CC metabolism and sequences further required for their regulation.  
 CC Epigenetic parameters are in particular cytosine methylations and  
 CC further chemical modifications of DNA bases of genes associated with  
 CC metabolism. Further epigenetic parameters include for e.g. the  
 CC acetylation of histones which correlates with DNA methylation.  
 CC AAS63306-AAS63373 represent chemically pretreated metabolism associated  
 CC genes, and related primers of the invention.  
 CC  
 XX Sequence 9265 BP; 2686 A; 133 C; 1801 G; 4645 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 242 Length: 9265  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x AAS63355 (1-9265)

QY 9 PheileilePheTrpIlePhe 15

DB 525 TTTATTATTTTGGATATT 545

## RESULT 12

ABL19746/c

ID ABL19746 standard; DNA; 15723 BP.

XX

AC ABL19746;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 10711.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

XX

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX

PS Claim 1; SEQ ID NO 10711; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 15723 BP; 5347 A; 2844 C; 2677 G; 4855 T; 0 other;

## Alignment Scores:

Pred. No.: 380 Length: 15723  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABL19746 (1-15723)

QY 9 PheileilePheTrpIlePhe 15

DB 13139 TTCATAATTTTGGATATT 13119

## RESULT 13

ABL70143

ID ABL70143 standard; DNA; 15881 BP.

XX

AC ABL70143;

XX

DT 01-JUL-2002 (first entry)

XX

DE Chemically treated cell signalling DNA sequence#17.

XX

KW Cell signalling; cytosine methylation; cell signalling disease; cancer; tumour; cytostatic; ds.

XX

OS Unidentified.

PN WO200202807-A2.

XX

PD 10-JAN-2002.

XX

PF 29-JUN-2001; 2001WO-EP07471.

XX

PR 30-JUN-2000; 2000DE-1032529.

XX

PR 01-SEP-2000; 2000DE-1043826.

XX

PA (EPIG-) EPIGENOMICS AG.

XX

PI Olek A, Piepenbrock C, Berlin K;

XX

DR WPI; 2002-154758/20.

XX

CC Nucleic acid, useful for diagnosis and therapy of diseases associated with cell signalling e.g. cancer, comprises chemically modified genomic sequences of genes associated with cell signalling -

XX

PS Claim 1; SEQ ID NO 33; 24pp+sequence listing; English.

XX

CC The invention relates to a nucleic acid comprising a sequence of at least 18 bases of a segment of chemically pretreated DNA of genes associated with cell signalling. The activity of the modified sequences of the invention may be described as cytostatic. The object of the invention is to provide the chemically modified DNA of genes associated with cell signalling, as well as oligonucleotides and/or RNA-oligomers for detecting cytosine methylations, as well as a method which is particularly suitable for the diagnosis and/or therapy of genetic and epigenetic parameters of genes associated with cell signalling. The chemically modified DNA provided by the invention is useful for diagnosis and therapy of diseases such as solid tumours and cancer. The sequences given in records ABL70111-ABL70626 represent chemically pre-treated genomic DNA's of genes associated with cell signalling.

CC Note: The sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

XX

SQ Sequence 15881 BP; 4515 A; 204 C; 3588 G; 7574 T; 0 other;

## Alignment Scores:

Pred. No.: 384 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABL70143 (1-15881)

QY 10 IleilePheTrpIlePheTrp 16

DB 11862 ATTATTTTGGATATTGG 11882

## RESULT 14

ABL32260

ID ABL32260 standard; DNA; 15881 BP.

XX

AC ABL32260;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 233.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytotatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200200928-A2.  
 PN  
 XX 03-JAN-2002.  
 PR  
 XX 02-JUL-2001; 2001WO-EP07537.  
 PF  
 XX 30-JUN-2000; 2000DE-1032529.  
 PR  
 XX 01-SEP-2000; 2000DE-1043826.  
 PD  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 XX WPI; 2002-130909/17.  
 DR  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation  
 PT  
 XX  
 PS Claim 1; SEQ ID NO 233; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 SQ Sequence 15881 BP; 4515 A; 204 C; 3588 G; 7574 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 384 Length: 15881  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-854-133-587 (1-16) x ABL32260 (1-15881)  
 QY 10 IlelePheTrpIlePheTrp 16  
 Db 11862 ATTATTTTGGATATTTGG 11882  
 RESULT 15  
 AAS61069  
 ID AAS61069 standard; DNA; 15881 BP.  
 XX  
 AC AAS61069;  
 XX  
 XX 29-JAN-2002 (first entry)  
 DT  
 XX Human gene regulation-associated gene oligonucleotide #24.  
 DE  
 XX

KW Human; Gene regulation-associated gene; severe combined immunodeficiency;  
 KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
 KW renal disease; Preclampsia; cardiac allograft vascular disease;  
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
 KW immunostimulant; cardiant; antinflammatory; coagulant; antiasthmatic;  
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200177375-A2.  
 PN  
 XX 18-OCT-2001.  
 PD  
 XX  
 XX 06-APR-2001; 2001WO-EP03968.  
 PF  
 XX 06-APR-2000; 2000DE-1019058.  
 PR  
 XX 07-APR-2000; 2000DE-1019173.  
 PR  
 XX 30-JUN-2000; 2000DE-1032529.  
 PR  
 XX 01-SEP-2000; 2000DE-1043826.  
 PD  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 XX WPI; 2002-017470/02.  
 DR  
 XX New nucleic acid sequences from chemically modified genes associated  
 PT with gene regulation, useful for analysing cytosine methylations for  
 PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
 PT disease  
 PT  
 XX  
 PS Claim 1; SEQ ID NO 25; 26pp; English.  
 XX  
 CC The invention relates to 224 nucleic acid sequences comprising at least  
 CC 18 bases of a chemically pretreated gene associated with gene regulation  
 CC selected from 43 known genes (or complementary sequences). The  
 CC chemical pretreatment converts cytosine bases unmethylated at the  
 CC 5-position to uracil or another base with hybridisation behaviour  
 CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
 CC The DNA sequences, oligomers (or sets/arrays) and method are  
 CC useful in the diagnosis of diseases (or predisposition to diseases)  
 CC associated with gene regulation and in therapy of such diseases, by  
 CC enabling analysis of the cytosine methylation patterns of such genes,  
 CC kits are provided. They are especially useful in diagnosis  
 CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
 CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
 CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease, a  
 CC preclampsia, graft versus-host disease. The present sequence is a  
 CC sequence included in the sequence data for this specification and is  
 CC associated with the human gene regulation-associated genes.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX  
 SQ Sequence 15881 BP; 4515 A; 204 C; 3588 G; 7574 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 384 Length: 15881  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-854-133-587 (1-16) x AAS61069 (1-15881)  
 QY 10 IlelePheTrpIlePheTrp 16  
 Db 11862 ATTATTTTGGATATTTGG 11882  
 RESULT 16



ABK31182  
ID ABK31182 standard; DNA; 15881 BP.

XX AC ABK31182;

XX DT 23-APR-2002 (first entry)

XX DE Signal transduction associated gene modified DNA #13.

XX KW Human; signal transduction associated gene; cytosine methylation state;  
KW Cpg island; signal transduction associated disease; solid tumour; cancer;  
KW antitumour; cytosine; mutant; ds.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200200926-A2.

XX PD 03-JAN-2002.

XX PF 29-JUN-2001; 2001WO-EP07472.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-147896/19.

XX PT Oligonucleotide for diagnosis and therapy of diseases associated with  
PT signal transduction e.g. cancer, comprises chemically modified genomic  
PT sequences of genes associated with signal transduction

XX PS Claim 1; SEQ ID No 25; 24pp; English.

XX CC The present invention relates to chemically modified DNA sequences of  
CC signal transduction associated genes. The DNA sequences are chemically  
CC modified using a solution of bisulphite, hydrogen sulphite or  
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers  
CC for detecting the cytosine methylation state (CpG islands) of these  
CC genes, and a method for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with signal transduction.  
CC The genomic DNA can be obtained from cells or cellular components which  
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,  
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from  
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,  
CC histologic object slides, and all their possible combinations. The  
CC sequences of the invention are useful for the diagnosis and therapy of  
CC diseases associated with signal transduction e.g. solid tumours and  
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA  
CC sequences of different genes associated with signal transduction, or  
CC their complementary sequences.

XX CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX SQ Sequence 15881 BP; 4515 A; 204 C; 3588 G; 7574 T; 0 other;

Alignment Scores:  
Pred. No.: 384 Length: 15881  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABK31182 (1-15881)

QY 10 llellePheTrpPheTrp 16  
|||||

Db 11862 ATTATTTTGGATATTTGG 11882

RESULT 17

ABN79984

ID ABN79984 standard; DNA; 16633 BP.

XX AC ABN79984;

XX DT 15-JUL-2002 (first entry)

XX DE Human chemically modified disease associated gene SEQ ID NO 1.

XX KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
KW antidiabetic; cytosine; anticonvulsant; ds.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200200927-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP07536.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-130908/17.

XX PT Novel nucleic acid useful for diagnosis and therapy of diseases  
PT associated with development genes such as diabetes, comprises a  
PT sequence of a segment of chemically pretreated DNA of genes associated  
PT with development

XX PS Claim 1; SEQ ID NO 1; 27pp; English.

XX CC The invention relates to a nucleic acid (I) comprising a sequence at  
CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
CC of genes associated with development selected from 87 genes listed in  
CC the specification such as ACCPN, ADFN, or AFD1 and comprising one of 350  
CC sequences (ABN79984-ABN80333) or their complements. The invention is  
CC useful for the diagnosis or therapy of diseases associated with  
CC development genes, in particular disease related to homeobox containing  
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
CC associated with congenital heart disease, epilepsy, diseases related to  
CC histone deacetylation, Currarino syndrome, diseases related with the  
CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
CC Oligomers specific to each of the genes are useful for detecting the  
CC methylation state of all Cpg dinucleotides within the 350 sequences or  
CC (II) and their complementary sequences, as primer oligonucleotides for  
CC the amplification of the 350 sequences, (II) and/or their complements and  
CC as oligomer probes for detecting the cytosine methylation state and/or  
CC single nucleotide polymorphisms (SNPs).

XX CC Note: The sequence data for this patent did not form part of the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.

XX SQ Sequence 16633 BP; 4452 A; 142 C; 3667 G; 8372 T; 0 other;

Alignment Scores:  
Pred. No.: 399 Length: 16633  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABN79984 (1-16633)

QY 9 PheillePheTTPilePhe 15  
ID AAV42717  
Db 11972 TTTATTTATTTTGGATTTC 11992

RESULT 18  
AAV42717  
AC AAV42717 standard; DNA; 26 BP.  
XX AAV42717;  
XX 14-OCT-1998 (first entry)  
DT 3' PCR primer used to amplify human beta galactosidase cDNA.  
DE Retroviral vector; gene delivery vehicle; expression; PCR primer;  
KW non-immunogenic selectable marker; gene therapy; activation;  
KW human; beta galactosidase; ss.  
XX Synthetic.  
OS Homo sapiens.  
XX WO9830709-A2.  
PN 16-JUL-1998.  
PD 14-JAN-1998; 98MO-US00715.  
PF 13-JAN-1998; 98US-0038339.  
PR 14-JAN-1997; 97US-0035473.  
PR 27-FEB-1997; 97US-0038339.  
XX (CHAD/) CHADA S.  
FA (JOLLY/) JOLLY D J.  
PA (MOOR/) MOORE M D.  
XX Chada S, Jolly DJ, Moore MD;  
PI WPI; 1998-399153/34.  
DR Non-immunogenic pro:drug activating enzyme(s) and selectable  
PT marker(s) are used in gene therapy for the treatment of a wide  
PT variety of disorder(s)  
XX Example 2; Page 46; 121pp; English.

XX PCR primers AAV42716-17 were used to amplify human beta galactosidase  
CC cDNA (see AAV42728). The amplified product is used as a genetic marker  
CC in the retroviral vector of the invention. The specification describes  
CC a gene delivery vehicle which directs expression of a non-immunogenic  
CC selectable marker or molecule which is capable of activating a  
CC previously inactive compound. Vectors expressing the markers and a  
CC heterologous sequence are useful in gene therapy. The vectors can be used  
CC to deliver a molecule into a target area where it may cause the  
CC activation of a previously inactive substance.  
XX Sequence 26 BP; 7 A; 5 C; 7 G; 7 T; 0 other;  
SQ Alignment Scores:  
Pred. No.: 19.9 Length: 26  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 19 Gaps: 0

US-09-854-133-587 (1-16) x AAV42717 (1-26)  
QY 6 GlyilleAspPheillelle 11  
Db 4 GGGATCGATTTCATC 21

RESULT 19

AAC30925/C  
ID AAC30925 standard; cDNA; 246 BP.  
XX AAC30925;  
AC AAC30925;  
XX 06-OCT-2000 (first entry)  
DT Human secreted protein 5' EST, SEQ ID NO: 35000.  
DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
KW Homo sapiens.  
XX EP1033401-A2.  
PN 06-SEP-2000.  
PD 21-FEB-2000; 2000EP-0200610.  
PF 26-FEB-1999; 99US-0122487.  
PR (GEST ) GENSET.  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
PI WPI; 2000-500381/45.  
DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 35000; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

SQ Sequence 246 BP; 74 A; 47 C; 55 G; 70 T; 0 other;  
Alignment Scores:  
Pred. No.: 136 Length: 246  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 21 Gaps: 0

US-09-854-133-587 (1-16) x AAC30925 (1-246)  
QY 10 IleillePheTTPilePhe 15  
Db 61 ATTATCTCTCGATTTC 44

RESULT 20  
ABL77073/C  
ID ABL77073 standard; cDNA; 250 BP.  
XX ABL77073;  
AC ABL77073;  
XX 17-MAY-2002 (first entry)  
DT XX

DE Human ovarian cancer related cDNA clone SEQ ID NO:51.  
 XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.  
 KW Homo sapiens.  
 OS WO200192581-A2.  
 PN 06-DEC-2001.  
 PD 29-MAY-2001; 2001WO-US17756.  
 XX 26-MAY-2000; 2000US-207484P.  
 PF (CORI-) CORIXA CORP.  
 PR Algate PA, Harlocker SL, Jones R;  
 PA WPI; 2002-122075/16.  
 XX Composition for therapy and diagnosis of ovarian cancer comprising  
 XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
 PT polypeptide, antibody specific to polypeptide or T cell expressing  
 PT polypeptide  
 XX  
 PS Claim 1; SEQ ID 51; 489pp; English.  
 XX The present invention describes a composition (I) comprising: carriers  
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour  
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
 CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to  
 CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
 CC population of (II), or antigen presenting cells that express (II).  
 CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to  
 CC (S1) can be used for detecting ovarian cancer in a patient's biological  
 CC sample preferably serum or ovarian tissue. The method comprises  
 CC contacting a biological sample from a patient with (IV), detecting the  
 CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
 CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
 CC patient, where the amount of polynucleotide hybridising to (IV) is  
 CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
 CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
 CC specific for an ovarian tumour protein comprising contacting T cells  
 CC with (III) or (II). (III) is useful in design and preparation of  
 CC ribozyme molecules for inhibiting expression of the tumour polypeptides  
 CC and proteins in tumour cells; and to isolate a full length gene from a  
 CC suitable library e.g., a tumour cDNA library using well known  
 CC techniques.  
 XX  
 SQ Sequence 250 BP; 106 A; 43 C; 27 G; 74 T; 0 other;

Alignment Scores:  
 Pred. No.: 138 Length: 250  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABL77073 (1-250)

QY 10 IleilePheTrpIlePhe 15  
 |||||  
 Db 100 ATAATTTTCGGATTTT 83

RESULT 21  
 AAS47046  
 ID AAS47046 standard; cDNA; 277 BP.  
 XX  
 AC AAS47046;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX

DE Human breast cancer cDNA clone 13082.

XX Human; ss; breast cancer protein; tumour; cancer; cytostatic;  
 KW gene therapy.  
 OS Homo sapiens.

PN WO200179286-A2.  
 PD 25-OCT-2001.

XX 12-APR-2001; 2001WO-US12164.  
 XX 17-APR-2000; 2000US-0551621.  
 PR 08-JUN-2000; 2000US-0590751.  
 PR 22-JUN-2000; 2000US-0604287.  
 PR 20-JUL-2000; 2000US-0620405.

XX (CORI-) CORIXA CORP.

XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;  
 PI WPI; 2001-611721/70.  
 XX Breast Tumour Proteins and nucleic acids useful for the prevention,  
 PT diagnosis and treatment of breast cancer  
 XX  
 PS Claim 1; Page 175; 297pp; English.

XX The invention relates to isolated breast tumour proteins and  
 CC nucleic acids that encode them, including immunogenic fragments of the  
 CC proteins. Also included are expression vectors expressing the  
 CC proteins, transformed cells and antibodies raised against the proteins or  
 CC an antigen presenting cell expressing the protein. The proteins and  
 CC nucleic acids may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate breast tumour protein expression,  
 CC i.e. breast tumours and breast cancer e.g by gene therapy. The nucleic  
 CC acids and their complements may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acids in  
 CC samples, and therefore which patients may be in need of restorative  
 CC therapy. The proteins, nucleic acids and antibodies may be used in assays  
 CC to identify modulators (e.g. antagonists) of breast tumour protein  
 CC expression and activity. The antibodies and antagonists may also be used  
 CC to down regulate expression and activity. The antibodies may also be used  
 CC as diagnostic agents for detecting the presence of the proteins in  
 CC samples (e.g. by enzyme linked immunosorbant assay (ELISA)) and in other  
 CC immuno-purification diagnostic techniques. The present sequence is  
 CC a cDNA from a breast tumour cDNA library isolated by subtractive  
 CC hybridisation against a normal breast cDNA library.

XX Sequence 277 BP; 97 A; 44 C; 80 G; 56 T; 0 other;

Alignment Scores:  
 Pred. No.: 151 Length: 277  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 22 Gaps: 0

US-09-854-133-587 (1-16) x AAS47046 (1-277)

QY 10 IleilePheTrpIlePhe 15  
 |||||  
 Db 96 ATTATTTTGGATCTTC 113

RESULT 22  
 AAF17616  
 ID AAF17616 standard; cDNA; 277 BP.  
 XX  
 AC AAF17616;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX

XX	Human breast cancer associated 13082 coding sequence.
DE	Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.
XX	Homo sapiens.
KW	WO200060076-A2.
XX	12-OCT-2000.
OS	15-FEB-2000; 2000WO-US05308.
PV	02-APR-1999; 99US-0285480.
XX	23-JUN-1999; 99US-0339338.
PR	02-SEP-1999; 99US-0389681.
PR	03-NOV-1999; 99US-0433826.
XX	(CORI-) CORIXA CORP.
PA	Yuqiu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;
XX	WPI; 2001-122627/13.
DR	An isolated polypeptide useful for the treatment and diagnosis of
PT	tumors e.g. breast cancer comprises at least an immunogenic portion of
PT	a breast tumor protein -
XX	Claim 6; Page 117; 238pp; English.
PS	The present invention provides the coding sequences and some protein
CC	sequences of proteins associated with breast cancer in humans. These
CC	sequences can be used in the diagnosis and treatment of cancers,
CC	particularly breast tumours.
XX	SQ Sequence 277 BP; 97 A; 44 C; 80 G; 56 T; 0 other;
XX	Alignment Scores:
Score:	151 Length: 277
Percent Similarity:	6.00 Matches: 6
Best Local Similarity:	100.00% Conservative: 0
Query Match:	100.00% Mismatches: 0
DB:	37.50% Indels: 0
	22 Gaps: 0
US-09-854-133-587 (1-16) x AAF17616 (1-277)	
QY	10 IlellePheTrpillePhe 15
DB	
	96 ATTATATTTCGAICTTC 113
RESULT 23	
ABK95081	
ID	ABK95081 standard; cDNA; 277 BP.
XX	AC ABK95081;
XX	24-SEP-2002 (first entry)
DT	Human breast tumour cDNA 13082.
XX	Human; ss; breast cancer; cytostatic; gene therapy; SYN22A12; SYN22A2;
DE	B723P; B726P.
XX	Homo sapiens.
OS	US6387697-B1.
PN	14-MAY-2002.
PD	28-DEC-1998; 98US-0222575.
PF	28-DEC-1998; 98US-0222575.
XX	PR

```

XX 26-MAY-2000; 2000US-207484P.
XX (CORI-) CORIXA CORP.
XX Algate PA, Harlocker SL, Jones R;
XX WPI; 2002-122075/16.
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
XX polypeptide, antibody specific to polypeptide or T cell expressing
XX polypeptide
XX
XX Claim 1; SEQ ID 10058; 489pp; English.
XX
XX The present invention describes a composition (I) comprising: carriers
XX and immunostimulants; and a polypeptide (II) of a ovarian tumour
XX polypeptide encoded by a polynucleotide (III) having a cDNA sequence
XX (S1) from the 10912 nucleotide sequences as given in ABL77023 to
XX ABL87934, (III) encoding (II) having a sequence (S2), a T cell
XX population of (II), or antigen presenting cells that express (II).
XX (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
XX (S1) can be used for detecting ovarian cancer in a patient's biological
XX sample preferably serum or ovarian tissue. The method comprises
XX contacting a biological sample from a patient with (IV), detecting the
XX amount of polynucleotide hybridising to (IV) and comparing the amount to
XX a predetermined cutoff value and thereby detecting ovarian cancer in the
XX patient, where the amount of polynucleotide hybridising to (IV) is
XX detected preferably by polymerase chain reaction (PCR). (I) comprising
XX (III) and/or (II) is useful for stimulating and/or expanding T cells
XX specific for an ovarian tumour protein comprising contacting T cells
XX with (III) or (II). (III) is useful in design and preparation of
XX ribozyme molecules for inhibiting expression of the tumour polypeptides
XX and proteins in tumour cells; and to isolate a full length gene from a
XX suitable library e.g., a tumour cDNA library using well known
XX techniques.
XX
XX Sequence 283 BP; 60 A; 71 C; 58 G; 94 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 153 Length: 283
XX Score: 6.00 Matches: 6
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 37.50% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-854-133-587 (1-16) x ABL87080 (1-283)
XX
XX QY 9 PheIleIlePheTrpIle 14
XX
XX Db 21 TTTATTATTTTGGATA 38
XX
XX RESULT 25
XX AAK77254/c
XX ID AAK77254 standard; DNA; 303 BP.
XX
XX AC AAK77254;
XX
XX DT 07-NOV-2001 (first entry)
XX
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32066.
XX
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200157182-A2.
XX
XX PD 09-AUG-2001.
XX
XX

```

```

PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

```

PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 01-NOV-2000; 2000US-0241826.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 03-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0254097.  
 PR 05-JAN-2001; 2001US-0256787.  
 (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI; 2001-483426/52.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,

PT useful for preventing, diagnosing and/or treating cancers and  
 metastasis -  
 XX  
 PS Disclosure; SEQ ID NO 32066; 3071pp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 303 BP; 98 A; 64 C; 64 G; 77 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 163 Length: 303  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-854-133-587 (1-16) x AAK77254 (1-303)  
 QY 7 IleasPhellellePhe 12  
 DB 71 ATAGATTTTATATCTTC 54  
 RESULT 26  
 ABV17400/C  
 ID ABV17400 standard; cDNA; 312 BP.  
 XX  
 AC ABV17400;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 17391.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US05171.  
 XX  
 PR 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 DR WPI; 2001-662795/76.  
 XX

PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX  
 PS Claim 1; Page 2881; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 312 BP; 94 A; 70 C; 53 G; 93 T; 2 other;  
 SQ

Alignment Scores:  
 Pred. No.: 167 Length: 312  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABV17400 (1-312)

QY 9 PheillelePheTrpIle 14

DB 67 TTCATTATTTTGGATT 50

RESULT 27

ABV47195/c

ID ABV47195 standard; cDNA; 339 BP.

XX

AC ABV47195;

XX

DT 16-SEP-2002 (first entry)

XX

DE Human prostate expression marker cDNA 47186.

XX

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX

KW pharmacogenomic marker; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200160860-A2.

XX

PD 23-AUG-2001.

XX

PF 20-FEB-2001; 2001WO-US05171.

XX

PR 17-FEB-2000; 2000US-183319P.

XX

PR 16-MAR-2000; 2000US-189862P.

XX

PR 25-MAY-2000; 2000US-207454P.

XX

PR 09-JUN-2000; 2000US-211314P.

XX

PR 18-JUL-2000; 2000US-219007P.

XX

PR 13-DEC-2000; 2000US-255281P.

XX

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Schlegel R, Endege WO, Monahan JE;

XX

DR WPI; 2001-662795/76.

XX

PT Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX  
 PS Claim 1; Page 9288; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 339 BP; 104 A; 81 C; 61 G; 93 T; 0 other;  
 SQ

Alignment Scores:  
 Pred. No.: 179 Length: 339  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABV47195 (1-339)

QY 9 PheillelePheTrpIle 14

DB 118 TTCATTATTTTGGATT 101

RESULT 28

AAS66468/c

ID AAS66468 standard; cDNA; 416 BP.

XX

AC AAS66468;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #2272.

XX

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

XX

DR P-PSDB; ABG02281.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX

PS Claim 1; SEQ ID NO 2272; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 416 BP; 75 A; 94 C; 61 G; 74 T; 112 other;

#### Alignment Scores:

Pred. No.:	213	Length:	416
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.50%	Indels:	0
DB:	23	Gaps:	0

US-09-854-133-587 (1-16) x AAS66468 (1-416)

QY 6 GlytLeAspPheIleIle 11  
 |||||

Db 338 GGCATTGATTTCATT 321

#### RESULT 29

AAV90289  
 ID AAV90289 standard; cDNA; 418 BP.

AC AAV90289;

XX 15-FEB-1999 (first entry)

DE EST clone DK70.

XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
 KW gene therapy; ss.

XX Homo sapiens.

XX WO9845436-A2.

PN 15-OCT-1998.

PF 10-APR-1998; 98WO-US06955.

PR 10-APR-1997; 97US-0838821.

XX (GEMY ) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Racie LA, Spaulding V, Treacy M;

XX WPI; 1999-070077/06.

PT New polynucleotides encoding human secreted proteins - derived from

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 XX ovary, pituitary, retina and colon cDNA libraries.  
 PS Claim 1; Page 492-493; 618pp; English.

XX The present sequence represents a human expressed sequence tag (EST).  
 CC The polynucleotide, which is a secreted EST, and the encoded protein  
 CC are predicted to have useful biological activities which would make  
 CC them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals, although no supporting data is  
 CC given. Suggested activities include nutritional activity, immune  
 CC stimulating or suppressing activity, haematopoiesis regulating  
 CC activity, tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
 CC activity, receptor/ligand activity, anti-inflammatory activity,  
 CC cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The polynucleotide may also be useful for gene therapy.  
 XX Sequence 418 BP; 130 A; 75 C; 99 G; 114 T; 0 other;

#### Alignment Scores:

Pred. No.:	214	Length:	418
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.50%	Indels:	0
DB:	20	Gaps:	0

US-09-854-133-587 (1-16) x AAV90289 (1-418)

QY 9 PheIleIlePheTrpIle 14  
 |||||

Db 381 TTTATAATATTGGATA 398

#### RESULT 30

ABN94582

ID ABN94582 standard; DNA; 498 BP.

XX AC ABN94582;

XX 13-AUG-2002 (first entry)

DE Gene #1080 used to diagnose liver cancer.

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
 KW metastatic liver tumour; cytostatic; expression profile; disease state;  
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

XX WO200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US30589.

XX 02-OCT-2000; 2000US-237054P.

XX (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer,  
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
 PT involves detecting the level of expression of two or more genes in a  
 PT liver tissue sample -

XX Claim 1; SEQ ID NO 1080; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the  
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver



CC tumour in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in RN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytostatic activity. The method is useful for diagnosing and detecting  
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 498 BP; 159 A; 114 C; 93 G; 132 T; 0 other;

Alignment Scores: 249 Length: 498  
 Pred. No.: 6.00 Matches: 6  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 37.50% Gaps: 0  
 DB: 24

US-09-854-133-587 (1-16) x ABN94582 (1-498)

QY 6 GlyIleAspPheIleIle 11  
 DB 48 GCATGACCTTCATTATA 65

RESULT 31  
 ABQ35704  
 ID ABQ35704 standard; DNA; 530 BP.

XX AC ABQ35704;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22295.  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

XX OS Homo sapiens.  
 XX WO200218632-A2.

XX PN 07-MAR-2002.

XX PD 01-SEP-2001; 2001WO-EP10074.

XX PF 01-SEP-2000; 2000DE-1043826.

XX PR 05-SEP-2000; 2000DE-1044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WIPI; 2002-371829/40.

XX DR Determining the degree of cytosine methylation in genomic DNA, useful  
 XX for diagnosis and prognosis, comprises selective hybridization of  
 XX PT amplicons from chemically treated DNA

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ34121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 530 BP; 95 A; 60 C; 146 G; 228 T; 1 other;

Alignment Scores: 263 Length: 530  
 Pred. No.: 6.00 Matches: 6  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 37.50% Gaps: 0  
 DB: 24

US-09-854-133-587 (1-16) x ABQ35704 (1-530)

QY 7 IleAspPheIleIlePhe 12

DB 190 ATAGATTATTATATTT 207

RESULT 32  
 ABQ35705/C  
 ID ABQ35705 standard; DNA; 530 BP.

XX AC ABQ35705;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22296.  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

XX OS Homo sapiens.  
 XX WO200218632-A2.

XX PN 07-MAR-2002.

XX PD 01-SEP-2001; 2001WO-EP10074.

XX PF 01-SEP-2000; 2000DE-1043826.

XX PR 05-SEP-2000; 2000DE-1044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WIPI; 2002-371829/40.

XX DR Determining the degree of cytosine methylation in genomic DNA, useful  
 XX for diagnosis and prognosis, comprises selective hybridization of  
 XX PT amplicons from chemically treated DNA

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised from the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ34121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

SQ Sequence 530 BP; 228 A; 146 C; 60 G; 95 T; 1 other;

Alignment Scores:  
 Pred. No.: 263 Length: 530  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABQ35705 (1-530)

QY 7 IleAspPheIlePhe 12

DB 341 ATAGATTTTATATATTT 324

RESULT 33

AAH1384

ID AAH11384 standard; cDNA; 548 BP.

AC AAH11384;

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer) SEQ ID NO:8219.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

PS Claim 3; SEQ ID 8219; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a polynucleotide which is complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

SQ Sequence 548 BP; 115 A; 161 C; 133 G; 130 T; 9 other;

Alignment Scores:

Pred. No.: 270 Length: 548  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 22 Gaps: 0

US-09-854-133-587 (1-16) x AAH11384 (1-548)

QY 1 PheGlnAlaAsnCysGly 6

DB 296 TTTCAGGCAATTCGGA 313

RESULT 34

ABQ45838

ID ABQ45838 standard; DNA; 553 BP.

AC ABQ45838;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32429.

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 553 BP; 105 A; 65 C; 190 G; 193 T; 0 other;

Alignment Scores: Pred. No.: 272 Length: 553  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABQ45838 (1-553)

QY 5 CysGlyIleAspPheIle 10  
 DB 251 TGGCGAATCGATTATT 268

RESULT 35  
 ABQ45839/C  
 ID ABQ45839 standard; DNA; 553 BP.  
 AC  
 AC ABQ45839;

12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 32430.  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

PN 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 553 BP; 193 A; 190 C; 65 G; 105 T; 0 other;

Alignment Scores: Pred. No.: 272 Length: 553  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABQ45839 (1-553)

QY 5 CysGlyIleAspPheIle 10

DB 303 TGGCGAATCGATTATT 286

RESULT 36

ABQ44134

ID ABQ44134 standard; DNA; 559 BP.

XX AC ABQ44134;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 30725.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX Sequence 559 BP; 67 A; 68 C; 216 G; 208 T; 0 other;

#### Alignment Scores:

Pred. No.:	275	Length:	559
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.50%	Indels:	0
DB:	24	Gaps:	0

US-09-854-133-587 (1-16) x ABQ44134 (1-559)

QY 8 AspPheIleIlePheTrp 13  
DB 517 GATTTTATTATTTTGG 534

RESULT 37  
ABQ44135/c

ID ABQ44135 standard; DNA; 559 BP.

XX ABQ44135;

AC 12-JUL-2002 (first entry)

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30726.  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.

XX Homo sapiens.

OS WO200218632-A2.

PN 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

PF 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

PA

XX

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

DR Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX Sequence 559 BP; 208 A; 216 C; 68 G; 67 T; 0 other;

#### Alignment Scores:

Pred. No.:	275	Length:	559
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.50%	Indels:	0
DB:	24	Gaps:	0

US-09-854-133-587 (1-16) x ABQ44135 (1-559)

QY 8 AspPheIleIlePheTrp 13

DB 43 GATTTTATTATTTTGG 26

RESULT 38

AAD05092/c

ID AAD05092 standard; cDNA; 572 BP.

XX AAD05092;

AC 17-JUL-2001 (first entry)

DT 17-JUL-2001 (first entry)

DE Human secreted protein-encoding gene 9 cDNA clone HDBAQ55, SEQ ID NO:50.

XX Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiotensin disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulnerability; gene therapy;  
KW cell culture; chemotaxis; food additive;  
KW binding partner identification; ss.

OS Homo sapiens.

XX

XX

PH

Location/Qualifiers

US-09-854-133-587 (1-16) x AAD05092 (1-572)

ID AAS85327 standard; cDNA: 582 bp.

us-09-854-133-587.oligo.rng

Tue May 13 12:12:58 2003

```

XX AC AAS85327;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #21131.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSEQ-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG21140.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 1; SEQ ID No 21131; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 582 BP; 170 A; 132 C; 158 G; 122 T; 0 other;

Alignment Scores:
Pred. No.: 284 Length: 582
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x AAS85327 (1-582)

QY 6 GlytAspPheIleIle 11
|||||
Db 136 GGTATTGACTTCATTATA 153

RESULT 42
ABA62551
ID ABA62551 standard; DNA; 600 BP.
XX ABA62551;
XX DT 01-FEB-2002 (first entry)

US-09-854-133-587 (1-16) x AAC08897 (1-584)

QY 6 GlytAspPheIleIle 11
|||||
Db 584 GGCATTGACTTCATTATA 567

RESULT 41
AAC08897/c
ID AAC08897 standard; cDNA; 584 BP.
XX AAC08897;
XX AC AAC08897;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 12972.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX PS Claim 1; SEQ ID 12972; 71pp + CD-ROM; English.
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
XX SQ Sequence 584 BP; 142 A; 112 C; 132 G; 198 T; 0 other;

Alignment Scores:
Pred. No.: 285 Length: 584
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 21 Gaps: 0

```

XX DE Human foetal liver single exon nucleic acid probe #10856.  
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157277-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00669.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX DR WPI; 2001-483447/52.  
 XX PS Human genome-derived single exon nucleic acid probes useful for  
 XX PS analyzing gene expression in human fetal liver -  
 XX PS Claim 1; SEQ ID NO 10856; 639pp + sequence listing; English.

XX CC The invention relates to a single exon nucleic acid probe for  
 XX CC measuring human gene expression in a sample derived from human foetal  
 XX CC liver. The single exon nucleic acid probes may be used for predicting,  
 XX CC measuring and displaying gene expression in samples derived from human  
 XX CC fetal liver. The present sequence is a single exon nucleic acid  
 XX CC probe of the invention.  
 XX CC Note: The sequence data for this patent did not form part of the  
 XX CC printed specification, but was obtained in electronic format directly  
 XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 600 BP; 165 A; 132 C; 122 G; 181 T; 0 other;

Alignment Scores:  
 Pred. No.: 292 Length: 600  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 22 Gaps: 0

US-09-854-133-587 (1-16) x ABA62551 (1-600)

QY 6 GlytLeaspPheille 11  
 Db 489 GGCATTGACTTCATCAT 506

RESULT 43

ABA29879  
 ID ABA29879 standard; DNA; 600 BP.  
 XX AC ABA29879;  
 XX DT 23-JAN-2002 (first entry)

DE Probe #8345 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;  
 XX KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 XX KW congenital heart disease; ss.  
 XX OS Homo sapiens.

XX PN WO200157274-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00666.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX DR WPI; 2001-488899/53.  
 XX PS Single exon nucleic acid probes for analyzing gene expression in human  
 XX PS hearts -  
 XX PS Claim 1; SEQ ID NO 8345; 530pp; English.

XX CC The present invention relates to single exon nucleic acid probes for  
 XX CC measuring human gene expression in a sample derived from human heart. The  
 XX CC present sequence is one such probe. The probes may be used for  
 XX CC predicting, measuring and displaying gene expression in samples derived  
 XX CC from the human heart via microarrays. By measuring gene expression, the  
 XX CC probes are useful for predicting, diagnosing, grading, staging,  
 XX CC monitoring and prognosing diseases of the human heart and vascular system  
 XX CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 XX CC congenital heart disease.  
 XX CC Note: The sequence data for this patent did not form part of the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 600 BP; 165 A; 132 C; 122 G; 181 T; 0 other;

Alignment Scores:  
 Pred. No.: 292 Length: 600  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 22 Gaps: 0

US-09-854-133-587 (1-16) x ABA29879 (1-600)

QY 6 GlytLeaspPheille 11  
 Db 489 GGCATTGACTTCATCAT 506

RESULT 44

AAK10907  
 ID AAK10907 standard; DNA; 600 BP.  
 XX AC AAK10907;  
 XX DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 10898.

XX Human; brain expressed exon; gene expression analysis; probe;  
 XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 XX KW epilepsy; cancer; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157275-A2.  
 XX SQ



```

PD 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00667.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX Example 4; SEQ ID NO: 10898; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX Sequence 600 BP; 165 A; 132 C; 122 G; 181 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 292 Length: 600
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 22 Gaps: 0
US-09-854-133-587 (1-16) x AAK10907 (1-600)
QY 6 GlyIleAspPheIleIle 11
DB 489 GGCATTGACTTCATCATT 506
RESULT 45
AAK36765
ID AAK36765 standard; DNA; 600 BP.
XX AAK36765;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed single exon probe SEQ ID NO: 11322.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00668.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.

```

```

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX Example 4; SEQ ID NO: 11322; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX Sequence 600 BP; 165 A; 132 C; 122 G; 181 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 292 Length: 600
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 22 Gaps: 0
US-09-854-133-587 (1-16) x AAK36765 (1-600)
QY 6 GlyIleAspPheIleIle 11
DB 489 GGCATTGACTTCATCATT 506
Search completed: May 11, 2003, 17:16:11
Job time : 45.2655 secs

```



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 16:16:15 ; Search time 8.92035 Seconds  
(without alignments)  
550.071 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FOANCGIDFIWFIFW 16

Scoring table:  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 873169

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters:

-O-/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_174133\_728/app\_query.fasta\_1.462  
-DB-issued Patents\_NA -OPWT=fastap -SUFFIX=oligo.rni -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=60 -DOCALLGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1\_1\_47@runat\_05052003\_174133\_728 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	37.5	277	4	US-09-222-575-101
2	6	37.5	1352	4	US-08-817-913-15
3	6	37.5	1470	4	US-09-134-001C-2404
4	6	37.5	1485	4	US-09-134-001C-1339
5	6	37.5	1734	4	US-08-817-913-16
6	6	37.5	1920	1	US-08-817-913-17
7	6	37.5	1969	1	US-07-721-761A-35
8	6	37.5	1969	4	US-07-978-687-35
9	6	37.5	1969	4	US-08-926-522-17
10	6	37.5	1969	5	PCT-US91-05801-35
11	6	37.5	2371	2	US-09-020-466-1
12	6	37.5	2371	3	US-09-192-659-1

13	6	37.5	2453	4	US-08-961-527-316
14	6	37.5	2682	4	US-09-020-465-1
15	6	37.5	3210	3	US-08-613-009A-6
16	6	37.5	3210	4	US-08-778-570B-6
17	6	37.5	3210	4	US-09-059-584-6
18	6	37.5	3334	4	US-09-255-984-1
19	6	37.5	3539	4	US-09-508-542-16
20	6	37.5	3660	3	US-08-613-009A-5
21	6	37.5	3660	4	US-08-778-570B-5
22	6	37.5	3660	4	US-09-059-584-5
23	6	37.5	4711	1	US-08-414-926A-1
24	6	37.5	4711	2	US-08-926-922-1
25	6	37.5	4711	3	US-09-253-682-1
26	6	37.5	5064	4	US-09-527-657-1
27	6	37.5	5064	4	US-08-936-107A-8
28	6	37.5	8554	4	US-08-961-527-98
29	6	37.5	10357	4	US-08-961-527-191
30	6	37.5	20303	1	US-08-370-975B-6
31	6	37.5	26764	1	US-08-370-975B-1
32	6	31.2	24	4	US-09-341-444A-28
33	5	31.2	48	4	US-08-985-659-2
34	5	31.2	66	3	US-09-046-247-138
35	5	31.2	117	1	US-08-450-834-4
36	5	31.2	222	4	US-09-134-001C-548
37	5	31.2	225	4	US-09-134-001C-2829
38	5	31.2	258	1	US-08-410-804-6
39	5	31.2	258	1	US-08-259-514-6
40	5	31.2	258	2	US-08-858-311-6
41	5	31.2	313	4	US-09-438-906-15
42	5	31.2	390	4	US-09-134-001C-2354
43	5	31.2	405	4	US-09-134-001C-2395
44	5	31.2	409	1	US-08-469-667-21
45	5	31.2	409	1	US-09-224-110-21
46	5	31.2	409	5	PCT-US95-07289-21
47	5	31.2	439	4	US-09-221-017B-162
48	5	31.2	520	4	US-08-961-527-344
49	5	31.2	583	4	US-09-221-017B-682
50	5	31.2	585	4	US-08-976-259-45
51	5	31.2	595	4	US-09-385-982-452
52	5	31.2	601	4	US-09-814-951A-13
53	5	31.2	611	4	US-09-376-113-4
54	5	31.2	624	4	US-09-221-017B-43
55	5	31.2	631	4	US-09-276-113-6
56	5	31.2	642	2	US-08-480-753-5
57	5	31.2	642	3	US-09-041-889-10
58	5	31.2	642	3	US-08-837-058-10
59	5	31.2	660	2	US-08-401-068-5
60	5	31.2	660	2	US-08-846-338-5

# ALIGNMENTS

RESULT 1  
US-09-222-575-101  
; Sequence 101, Application US/09222575  
; Patent No. 6387697  
; GENERAL INFORMATION:  
; APPLICANT: Yuqiu, Jiang  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer  
; FILE REFERENCE: 210121.470  
; CURRENT APPLICATION NUMBER: US/09/222.575  
; CURRENT FILING DATE: 1998-12-28  
; NUMBER OF SEQ ID NOS: 174  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 101  
; LENGTH: 277  
; TYPE: DNA  
; ORGANISM: Human  
US-09-222-575-101

```

Alignment Scores:
Pred. No.: 20.1 Length: 277
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-222-575-101 (1-277)

QY 10 lletlePheTtPlePhe 15
Db 96 ATTATATTGGATCTTC 113

RESULT 2
US-08-817-913-15
; Sequence 15, Application US/08817913,
; Patent No. 6184443
; GENERAL INFORMATION:
; APPLICANT: Pedersen, Rolf
; APPLICANT: Lund, Marianne
; APPLICANT: Okkels, Finn
; APPLICANT: Kreiberg, Jette
; TITLE OF INVENTION: PROMOTER SEQUENCE FROM POTATO
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobb, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,913
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/02196
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: GB941286.7
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY0010.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1352 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-817-913-15

Alignment Scores:
Pred. No.: 79.5 Length: 1352
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-08-817-913-15 (1-1352)

```

```

QY 7 lleafpPheillePhe 12
Db 41 ATAGACTTCATAAATTTT 58

RESULT 3
US-09-134-001C-2404
; Sequence 2404, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2404
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2404

Alignment Scores:
Pred. No.: 85.5 Length: 1470
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-134-001C-2404 (1-1470)

QY 8 AspPheillePheTtP 13
Db 987 GATTTATATATCTGG 1004

RESULT 4
US-09-134-001C-1339/C
; Sequence 1339, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1339
; LENGTH: 1485
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1339

Alignment Scores:
Pred. No.: 86.3 Length: 1485
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-134-001C-1339 (1-1485)

```

```

QY 7 Ileasppheillellephe 12
DB 402 ATCGATTTCATATTTTC 385

RESULT 5
US-08-817-913-16
; Sequence 16, Application US/08817913
; Patent No. 6184443
; GENERAL INFORMATION:
; APPLICANT: Pedersen, Rolf
; APPLICANT: Lund, Marianne
; APPLICANT: Okkels, Finn
; APPLICANT: Kreiberg, Jette
; TITLE OF INVENTION: PROMOTER SEQUENCE FROM POTATO
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,913
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/02196
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: GB941286.7
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E.
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY0010.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1734 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-817-913-16

Alignment Scores:
Pred. No.: 98.7 Length: 1734
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-08-817-913-16 (1-1734)

QY 7 Ileasppheillellephe 12
DB 240 ATAGACTTCATAATTTT 257

RESULT 6
US-08-817-913-17
; Sequence 17, Application US/08817913
; Patent No. 6184443
; GENERAL INFORMATION:
; APPLICANT: Pedersen, Rolf

```

```

; APPLICANT: Lund, Marianne
; APPLICANT: Okkels, Finn
; APPLICANT: Kreiberg, Jette
; TITLE OF INVENTION: PROMOTER SEQUENCE FROM POTATO
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,913
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/02196
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: GB941286.7
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E.
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY0010.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-817-913-17

Alignment Scores:
Pred. No.: 108 Length: 1920
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-08-817-913-17 (1-1920)

QY 7 Ileasppheillellephe 12
DB 240 ATAGACTTCATAATTTT 257

RESULT 7
US-07-721-761A-35
; Sequence 35, Application US/07721761A
; Patent No. 5475099
; GENERAL INFORMATION:
; APPLICANT: Vic. C. Knauf
; APPLICANT: Gregory A. Thompson
; TITLE OF INVENTION: Plant Fatty Acid Synthases
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/721,761A
; FILING DATE: 19910626
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/568,493
; FILING DATE: 15-AUGUST-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; REFERENCE/DOCKET NUMBER: CGNE 76-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1969 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-07-721-761A-35
;
; Alignment Scores:
; Pred. No.: 110 Length: 1969
; Score: 6.00 Matches: 6
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 37.50% Indels: 0
; DB: 1 Gaps: 0
;
; US-09-854-133-587 (1-16) x US-07-721-761A-35 (1-1969)
;
; Qy 9 PheillelePheTrpIle 14
; Db 160 TTTATCATCTCTCTGGATC 177
;
; RESULT 8
; US-07-978-687-35
; Sequence 35, Application US/07978687
; Patent No. 5510255
; GENERAL INFORMATION:
; APPLICANT: Vic. C. Knauf
; APPLICANT: Gregory A. Thompson
; TITLE OF INVENTION: Plant Fatty Acid Synthases
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/978,687
; FILING DATE: FEBRUARY 1, 1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05801
; FILING DATE: 15-AUGUST-1991
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 07/568,493
; FILING DATE: 15-AUGUST-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/721,761
; FILING DATE: 26-JUNE-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; REFERENCE/DOCKET NUMBER: CGNE 76-2 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1969 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-07-978-687-35
;
; Alignment Scores:
; Pred. No.: 110 Length: 1969
; Score: 6.00 Matches: 6
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 37.50% Indels: 0
; DB: 1 Gaps: 0
;
; US-09-854-133-587 (1-16) x US-07-978-687-35 (1-1969)
;
; Qy 9 PheillelePheTrpIle 14
; Db 160 TTTATCATCTCTCTGGATC 177
;
; RESULT 9
; US-08-926-522-17
; Sequence 17, Application US/08926522
; Patent No. 6426447
; GENERAL INFORMATION:
; APPLICANT: Vic. C. Knauf
; APPLICANT: Gregory A. Thompson
; TITLE OF INVENTION: PLANT SEED OILS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1(a)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/458,173
; FILING DATE: 2-June-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE DES

```

us-09-854-133-587.oligo.rni

The May 13 12:12:58 2003

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1969 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mrna
US-08-926-522-17
Alignment Scores:
Pred. No.: 110 Length: 1969
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 5 Gaps: 0

US-09-854-133-587 (1-16) x PCT-US91-05801-35 (1-1969)
QY 9 PheillelePheTriple 14
Db 160 TTATCATCTTCTGGATC 177

RESULT 11
US-09-020-466-1
; Sequence 1, Application US/09020466
; Patent No. 5879908
; GENERAL INFORMATION:
; APPLICANT: LAPING, NICHOLAS
; APPLICANT: ZHU, YUAN
; TITLE OF INVENTION: CRFG-1a, a target and marker
; TITLE OF INVENTION: for chronic renal failure
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,466
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/045,203
; FILING DATE: 30-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70009-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2371 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-020-466-1
Alignment Scores:
Pred. No.: 130 Length: 2371
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 2 Gaps: 0

US-09-854-133-587 (1-16) x US-09-020-466-1 (1-2371)
QY 10 lleillePheTriplePhe 15

```

```

Db 1226 ATTATTTGGATCTTC 1243
RESULT 12
US-09-192-659-1
; Sequence 1, Application US/09192659
; Patent No. 6127522
; GENERAL INFORMATION:
; APPLICANT: LAPING, NICHOLAS J.
; APPLICANT: ZHU, YUAN
; TITLE OF INVENTION: CRFG-1a, A TARGET AND MARKER
; TITLE OF INVENTION: FOR CHRONIC RENAL FAILURE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,659
; FILING DATE: 16-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/020,466
; FILING DATE: 09-FEB-1998
; APPLICATION NUMBER: 60/045,203
; FILING DATE: 30-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70009-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2371 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-09-192-659-1
Alignment Scores:
Pred. No.: 130 Length: 2371
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 3 Gaps: 0
US-09-854-133-587 (1-16) x US-09-192-659-1 (1-2371)
QY 10 IlellePheTrpIlePhe 15
|||||
Db 1226 ATTATTTGGATCTTC 1243
RESULT 13
US-08-961-527-316
; Sequence 316, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

```

```

; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340Pl
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 316:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2453 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-316
Alignment Scores:
Pred. No.: 133 Length: 2453
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 4 Gaps: 0
US-09-854-133-587 (1-16) x US-08-961-527-316 (1-2453)
QY 9 PheIlellePheTrpIle 14
|||||
Db 940 TTTATTATCTTTGGATA 957
|||||
RESULT 14
US-09-020-465-1
; Sequence 1, Application US/09020465
; Patent No. 6255471
; GENERAL INFORMATION:
; APPLICANT: LAPING, NICHOLAS
; APPLICANT: OLSON, BARBARA
; APPLICANT: ZHU, YUAN
; TITLE OF INVENTION: CRFG-1b, a target and marker
; TITLE OF INVENTION: for chronic renal failure
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:

```



APPLICATION NUMBER: US/09/020,465  
FILING DATE: 09-FEB-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA: 60/045,203  
FILING DATE: 30-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70009-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2682 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-09-020-465-1

Alignment Scores:  
Pred. No.: 144 Length: 2682  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-020-465-1 (1-2682)

QY 10 IlelePheTrpIlePhe 15  
|||||  
Db 1537 ATTATATTGGATCTTC 1554

RESULT 15  
US-08-613-009A-6/c  
; Sequence 6, Application US/08613009A  
; Patent No. 6090576  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/613.009A  
; FILING DATE: 08-MAR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-542  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3210 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-613-009A-6  
Alignment Scores:  
Pred. No.: 169 Length: 3210  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x US-08-613-009A-6 (1-3210)

QY 11 IlePheTrpIlePheTrp 16  
|||||  
Db 28 ATTTTGGATTTTGG 11

RESULT 16  
US-08-778-570B-6/c  
; Sequence 6, Application US/08778570B  
; Patent No. 6437096  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/778.570B  
; FILING DATE: 03-JAN-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-664  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3210 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-778-570B-6  
Alignment Scores:  
Pred. No.: 169 Length: 3210  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

FILE REFERENCE: 10496/P65266U50  
CURRENT APPLICATION NUMBER: US/09/508,542  
CURRENT FILING DATE: 2000-05-16  
PRIORITY APPLICATION NUMBER: PCT/EP98/06136  
PRIORITY FILING DATE: 1998-09-26  
PRIORITY APPLICATION NUMBER: 97 116 841.4  
PRIORITY FILING DATE: 1997-09-27

;; PRIOR APPLICATION NUMBER: 97 118 755.4  
;; PRIOR FILING DATE: 1997-10-29  
;; NUMBER OF SEQ ID NOS: 23  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 16  
;; LENGTH: 3539  
;; TYPE: DNA  
;; ORGANISM: Staphylococcus carnosus  
;; FEATURE:  
;; NAME/KEY: modified\_base  
;; LOCATION: (1)..(3539)  
;; OTHER INFORMATION: "n" represents a, t, c, g, unknown or other  
US-09-508-542-16

Alignment Scores: 183 Length: 3539  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-508-542-16 (1-3539)

QY 6 GlylleasPherlelle 11  
|||||

Db 565 GGGATTGACTTTATTATC 582

RESULT 20

US-08-613-009A-5/c  
; Sequence 5, Application US/08613009A  
; Patent No. 6090576  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/613,009A  
FILING DATE: 08-MAR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-542  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3660 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-613-009A-5

Alignment Scores: 189 Length: 3660  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x US-08-613-009A-5 (1-3660)

QY 11 IlePheTrpIlePheTrp 16  
|||||

Db 295 ATTTTGGATTTTTGG 278

RESULT 21

US-08-778-570B-5/c  
; Sequence 5, Application US/08778570B  
; Patent No. 6437096  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/778,570B  
FILING DATE: 03-JAN-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-664  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3660 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-778-570B-5

Alignment Scores: 189 Length: 3660  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-08-778-570B-5 (1-3660)

QY 11 IlePheTrpIlePheTrp 16  
|||||

Db 295 ATTTTGGATTTTTGG 278

RESULT 22  
 US-09-059-584-5/c  
 ; Sequence 5, Application US/09059584  
 ; Patent No. 6440701  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Myers, Lisa E  
 ; APPLICANT: Schryvers, Anthony B  
 ; APPLICANT: Harkness, Robin E  
 ; APPLICANT: Loosmore, Sheena M.  
 ; APPLICANT: Du, Run-Pan  
 ; APPLICANT: Yang, Yan-Ping  
 ; APPLICANT: Klein, Michel H  
 ; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
 ; NUMBER OF SEQUENCES: 60  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Slim & McBurney  
 ; STREET: 6th Floor, 330 University Avenue  
 ; CITY: Toronto  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: M5G 1R7  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/059,584  
 ; FILING DATE: 14-APR-1998  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/778,570  
 ; FILING DATE: 03-JAN-1997  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Stewart, Michael I  
 ; REGISTRATION NUMBER: 24973  
 ; REFERENCE/DOCKET NUMBER: 1038-794  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (416) 595-1155  
 ; TELEFAX: (416) 595-1163  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3660 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-09-059-584-5

Alignment Scores:  
 Pred. No.: 189 Length: 3660  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-059-584-5 (1-3660)  
 QY 11 IlePheTrpIlePheTrp 16  
 Db 295 ATTTTGGATTGG 278

RESULT 23  
 US-08-414-926A-1/c  
 ; Sequence 1, Application US/08414926A  
 ; Patent No. 5721354  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Spaete, Richard  
 ; APPLICANT: Cha, Tai-An  
 ; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
 ; NUMBER OF SEQUENCES: 27  
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
 ; STREET: 5 Palo Alto Square  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94306-2155  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/414,926A  
 ; FILING DATE: March 31, 1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Cseri, Luann  
 ; REGISTRATION NUMBER: 31,822  
 ; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-494-7622  
 ; TELEFAX: 415-857-0663  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4711 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Human CMV  
 ; STRAIN: Towne  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: complement (845..1321)  
 ; OTHER INFORMATION: /product="UL147"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: complement (1368..1721)  
 ; OTHER INFORMATION: /product="UL152"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: complement (2504..3337)  
 ; OTHER INFORMATION: /product="UL153"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: complement (3515..4711)  
 ; OTHER INFORMATION: /product="UL154"  
 ; US-08-414-926A-1

Alignment Scores:  
 Pred. No.: 235 Length: 4711  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.50% Indels: 0  
 DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x US-08-414-926A-1 (1-4711)  
 QY 9 PheIleIlePheTrpIle 14  
 Db 4654 TTTATTATTTTGGATT 4637

RESULT 24  
 US-08-926-922-1/c  
 ; Sequence 1, Application US/08926922  
 ; Patent No. 5925751  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Spaete, Richard  
 ; APPLICANT: Cha, Tai-An

```

; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,922
; FILING DATE: September 10, 1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human CMV
; STRAIN: Towne
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (845..1321)
; OTHER INFORMATION: /product= "UL147"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (1368..1721)
; OTHER INFORMATION: /product= "UL152"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (2504..3337)
; OTHER INFORMATION: /product= "UL153"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (3515..4711)
; OTHER INFORMATION: /product= "UL154"
; US-08-926-922-1
;
Alignment Scores:
Pred No.: 235 Length: 4711
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 2 Gaps: 0

US-09-854-133-587 (1-16) x US-08-926-922-1 (1-4711)
Qy 9 PheillePheTrpIle 14
Db 4654 TTTATTATTTTGGATT 4637

RESULT 25
US-09-253-682-1/c
; Sequence 1, Application US/09253682
; Patent No. 6040170

```

```

; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/253,682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/926,922
; FILING DATE: September 10, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human CMV
; STRAIN: Towne
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (845..1321)
; OTHER INFORMATION: /product= "UL147"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (1368..1721)
; OTHER INFORMATION: /product= "UL152"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (2504..3337)
; OTHER INFORMATION: /product= "UL153"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (3515..4711)
; OTHER INFORMATION: /product= "UL154"
; US-09-253-682-1
;
Alignment Scores:
Pred No.: 235 Length: 4711
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x US-09-253-682-1 (1-4711)
Qy 9 PheillePheTrpIle 14
Db 4654 TTTATTATTTTGGATT 4637

```

Db 4654 TTTATTATTTTGGATT 4637

## RESULT 26

US-09-527-657-1/c

; Sequence 1, Application US/09527657

; Patent No. 6291236

; GENERAL INFORMATION:

; APPLICANT: Saete, Richard

; Cha, Tai-An

; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Luann Cseri Attorney at Law

; STREET: 750 Arimo Avenue

; CITY: Oakland

; STATE: CA

; COUNTRY: USA

; ZIP: 94610

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/527,657

; FILING DATE: 17-Mar-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/926,922

; FILING DATE: September 10, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Cseri, Luann

; REGISTRATION NUMBER: 31,822

; REFERENCE/DOCKET NUMBER: AVIR 11A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 510-834-1448

; TELEFAX: 510-839-7810

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4711 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Human CMV

; STRAIN: Towne

; FEATURE:

; NAME/KEY: CDS

; LOCATION: complement (845..1321)

; OTHER INFORMATION: /product= "UL147"

; FEATURE:

; NAME/KEY: CDS

; LOCATION: complement (1368..1721)

; OTHER INFORMATION: /product= "UL152"

; FEATURE:

; NAME/KEY: CDS

; LOCATION: complement (2504..3337)

; OTHER INFORMATION: /product= "UL153"

; FEATURE:

; NAME/KEY: CDS

; LOCATION: complement (3515..4711)

; OTHER INFORMATION: /product= "UL154"

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-527-657-1

## Alignment Scores:

Pred. No.: 235 Length: 4711

Score: 6.00 Matches: 6

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-527-657-1 (1-4711)

QY 9 Pheillelephetripile 14

|||||

Db 4654 TTTATTATTTTGGATT 4637

## RESULT 27

US-08-936-107A-8/c

; Sequence 8, Application US/08936107A

; Patent No. 6403306

; GENERAL INFORMATION:

; APPLICANT: Stephens, David S.

; TITLE OF INVENTION: Serogroup-Specific Nucleotide Sequences

; TITLE OF INVENTION: in the Molecular Typing of Bacterial Isolates and the

; NUMBER OF SEQUENCES: 51

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

; STREET: 5370 Manhattan Circle, Suite 201

; CITY: Boulder

; STATE: Colorado

; COUNTRY: US

; ZIP: 80303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/936,107A

; FILING DATE: 23-SEP-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/827,622

; FILING DATE: 09-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Caruthers, Jennie M.

; REGISTRATION NUMBER: 34,464

; REFERENCE/DOCKET NUMBER: 77-97

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 499-8080

; TELEFAX: (303) 499-8089

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5064 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: not relevant

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 479..1597

; NAME/KEY: CDS

; LOCATION: 1599..3236

; NAME/KEY: CDS

; LOCATION: 3309..4052

; NAME/KEY: CDS

; LOCATION: 4054..4917

US-08-936-107A-8

## Alignment Scores:

Pred. No.: 250 Length: 5064

Score: 6.00 Matches: 6

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 37.50% Indels: 0  
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-08-936-107A-8 (1-5064)

QY 10 llellePheTrpIlePhe 15  
|||||

Db 3530 ATCATCTTCTGTGATATT 3513

## RESULT 28

US-08-961-527-98/c  
; Sequence 98, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 98:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8654 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-961-527-98

Alignment Scores: 399 Length: 8654  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB: 4

US-09-854-133-587 (1-16) x US-08-961-527-98 (1-8654)

QY 5 CysGlyIleAspPheIle 10  
|||||

Db 233 TGGCGCATTTGATTCATC 216

## RESULT 29

US-08-961-527-191

; Sequence 191, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 191:

SEQUENCE CHARACTERISTICS:

LENGTH: 10357 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-191

Alignment Scores: 466 Length: 10357  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB: 4

US-09-854-133-587 (1-16) x US-08-961-527-191 (1-10357)

QY 7 IleAspPheIleIlePhe 12  
|||||

Db 9885 ATTGATTATATATCTTT 9902

## RESULT 30

US-08-370-975B-6/c

; Sequence 6, Application US/08370975B

; Patent No. 5622851

; GENERAL INFORMATION:

; APPLICANT: Maley, Frank

; APPLICANT: Maley, Gladys F.

; APPLICANT: Weiner, Karen X.B.

; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon, Hargrave, Devans & Doyle

; STREET: Clinton Square, P.O. Box 1051

; CITY: Rochester

; STATE: New York

; COUNTRY: USA

; ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/370,975B

FILING DATE: 10-JAN-1995

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1600
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-6

Alignment Scores:
Pred. No.: 836 Length: 20303
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x US-08-370-975B-6 (1-20303)

OY 10 llellepHeTrpIlePhe 15
Db 4242 ATTATCTTTGGATTTT 4225

RESULT 31
US-08-370-975B-1/c
; Sequence 1, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1600
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26764 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-1

Alignment Scores:
Pred. No.: 1,06e+03 Length: 26764
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 1 Gaps: 0

US-09-854-133-587 (1-16) x US-08-370-975B-1 (1-26764)

OY 10 llellepHeTrpIlePhe 15
Db 6205 ATTATCTTTGGATTTT 6188

RESULT 32
US-09-341-444A-28
; Sequence 28, Application US/09341444A
; Patent No. 6440866
; GENERAL INFORMATION:
; APPLICANT: Groenen, Martinus Antonius Mathilda
; APPLICANT: Albers, Gerardus Antonius Arnoldus
; TITLE OF INVENTION: Selection For Dwarfism in Poultry
; FILE REFERENCE: 310-1009
; CURRENT APPLICATION NUMBER: US/09/341,444A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/NL98/00021
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: EP 97200070.7
; PRIOR FILING DATE: 1997-01-10
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
; LENGTH: 24
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: OTHER INFORMATION: Oligonucleotide primer
US-09-341-444A-28

Alignment Scores:
Pred. No.: 33.4 Length: 24
Score: 5.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.25% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-341-444A-28 (1-24)

OY 7 lleAsppHeIleIle 11
Db 6 ATTGATTTCATCAT 20

RESULT 33
US-08-985-659-2/c
; Sequence 2, Application US/08985659
; Patent No. 6211134
; GENERAL INFORMATION:
; APPLICANT: Caldwell, Robert M
; APPLICANT: Mitchinson, Colin
; APPLICANT: Ropp, Traci M
; TITLE OF INVENTION: Mutant Alpha-Amylase
; FILE REFERENCE: A-69396/DJB/DAV/JJD
; CURRENT APPLICATION NUMBER: US/08/985,659
; CURRENT FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 08/645,971
; PRIOR FILING DATE: 1996-05-14
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
```



; SEQ ID NO 2  
; LENGTH: 48  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-08-985-659-2

Alignment Scores: 61 Length: 48  
Pred. No.: 5.00 Matches: 5  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 31.25% Gaps: 0  
DB: 4

US-09-854-133-587 (1-16) x US-08-985-659-2 (1-48)

QY 9 PheillelePheTrp 13  
Db 45 TTTATTATTTTGG 31

RESULT 34  
US-09-046-247-138  
; Sequence 138, Application US/09046247  
; Patent No. 612449  
; GENERAL INFORMATION:  
; APPLICANT: NIKOS PAGRATIS  
; APPLICANT: LARRY GOLD  
; TITLE OF INVENTION: HIGH AFFINITY TGF? NUCLEIC  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Swanson and Bratschun, L.L.C.  
; STREET: 8400 East Prentice Avenue, Suite #200  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: Word 7.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/046,247  
; FILING DATE: 23-MARCH-1998  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/458,424  
; FILING DATE: 2-JUNE-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/714,131  
; FILING DATE: 10-JUNE-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/536,428  
; FILING DATE: 11-JUNE-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/964,624  
; FILING DATE: 21-OCTOBER-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/117,991  
; FILING DATE: 8-SEPTEMBER-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/931,473  
; FILING DATE: 17-AUGUST-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barry Swanson  
; REGISTRATION NUMBER: 33,215  
; REFERENCE/DOCKET NUMBER: NEX 34.2/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 793-3333  
; TELEFAX: (303) 793-3433

; INFORMATION FOR SEQ ID NO: 138:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 66 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: RNA  
; FEATURE:  
; OTHER INFORMATION: All pyrimidines are 2'-F modified  
US-09-046-247-138

Alignment Scores: 80.4 Length: 66  
Pred. No.: 5.00 Matches: 5  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 31.25% Gaps: 0  
DB: 3

US-09-854-133-587 (1-16) x US-09-046-247-138 (1-66)

QY 5 CysGlyIleaspPhe 9  
Db 11 UGCGGUUUGACUUU 25

RESULT 35  
US-08-450-834-4  
; Sequence 4, Application US/08450834  
; Patent No. 5773705  
; GENERAL INFORMATION:  
; APPLICANT: Vierstra, Richard D  
; APPLICANT: Hondred, David  
; APPLICANT: Callis, Judy  
; TITLE OF INVENTION: Ubiquitin Fusion Protein System for  
; TITLE OF INVENTION: Protein Production in Plants  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: P.O. Box 2113  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/450,834  
; FILING DATE: 25-MAY-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/999,709  
; FILING DATE: 31-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.92425  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: UBQ-GUS

```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4..9
; OTHER INFORMATION: /function= "Sac II restriction site"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 39..44
; OTHER INFORMATION: /function= "Xho I restriction site"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 103..108
; OTHER INFORMATION: /function= "Bcl I restriction site"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 109..114
; OTHER INFORMATION: /function= "Xba I restriction site"
;
; US-08-450-834-4
```

```
Alignment Scores:
Pred. No.: 132 Length: 117
Score: 5.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.25% Indels: 0
DB: 1 Gaps: 0
```

US-09-854-133-587 (1-16) x US-08-450-834-4 (1-117)

```
QY 4 AsnCysGlyTleAsp 8
|||||
DB 92 AACTGTGGATTCAT 106
```

## RESULT 36

```
US-09-134-001C-548/c
; Sequence 548, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
```

```
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 548
; LENGTH: 222
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
```

```
US-09-134-001C-548
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,804
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1389
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```
Alignment Scores:
Pred. No.: 231 Length: 222
Score: 5.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.25% Indels: 0
DB: 4 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-134-001C-548 (1-222)

```
QY 11 IlePheTrpIlePhe 15
|||||
DB 164 ATTTCTGGATCTC 150
```

## RESULT 37

```
US-09-134-001C-2829
; Sequence 2829, Application US/09134001C
; Patent No. 6380370
```

```
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2829
; LENGTH: 225
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-2829
```

```
Alignment Scores:
Pred. No.: 233 Length: 225
Score: 5.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.25% Indels: 0
DB: 4 Gaps: 0
```

US-09-854-133-587 (1-16) x US-09-134-001C-2829 (1-225)

```
QY 9 PheIleIlePheTrp 13
|||||
DB 118 TTTATAATTTTGG 132
```

## RESULT 38

```
US-08-410-804-6
; Sequence 6, Application US/08410804
; Patent No. 5632994
; GENERAL INFORMATION:
```

```
; APPLICANT: Reed, John C.
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive, Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,804
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1389
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

Tue May 13 12:12:58 2003

us-09-854-133-587.oligo.rni

```
; MOLECULE TYPE: CDNA
US-08-410-804-6

Alignment Scores:
Pred. No.: 263
Score: 5.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 31.25%
Indels: 0
Gaps: 0
DB: 1

US-09-854-133-587 (1-16) x US-08-410-804-6 (1-258)

QY 7 IleAspPheillelle 11
Db 172 ATAGATTCATAATT 186

RESULT 39
US-08-259-514-6
; Sequence 6, Application US/08259514
; Patent No. 5747245
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive. Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/259,514
; FILING DATE: 14-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/410,804
; FILING DATE: 27-MAR-1995
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1389
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-259-514-6

Alignment Scores:
Pred. No.: 263
Score: 5.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 31.25%
Indels: 0
Gaps: 0
DB: 1

US-09-854-133-587 (1-16) x US-08-259-514-6 (1-258)

QY 7 IleAspPheillelle 11
Db 172 ATAGATTCATAATT 186

RESULT 40
US-08-259-514-6
; Sequence 6, Application US/08259514
; Patent No. 5747245
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive. Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/259,514
; FILING DATE: 14-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/410,804
; FILING DATE: 27-MAR-1995
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9954
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-259-514-6

Alignment Scores:
Pred. No.: 263
Score: 5.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 31.25%
Indels: 0
Gaps: 0
DB: 1

US-09-854-133-587 (1-16) x US-08-259-514-6 (1-258)

QY 7 IleAspPheillelle 11
Db 172 ATAGATTCATAATT 186

RESULT 41
US-09-438-906-15/C
; Sequence 15, Application US/09438906
; Patent No. 6465185
; GENERAL INFORMATION:
; APPLICANT: Goldfine, Ira
; APPLICANT: Trischitta, Vincenzo
; APPLICANT: Pizzuti, Antonio
; APPLICANT: Vigneri, Riccardo
; APPLICANT: Frittitta, Lucia
; TITLE OF INVENTION: Polymorphic Human PC-1 Sequences
; FILE REFERENCE: 9076-089
; CURRENT APPLICATION NUMBER: US/09/438,906
; CURRENT FILING DATE: 1999-11-18
```

```
US-08-858-311-6
; Sequence 6, Application US/08858311
; Patent No. 5876939
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive. Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,311
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/410,804
; FILING DATE: 27-MAR-1995
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1389
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-858-311-6

Alignment Scores:
Pred. No.: 263
Score: 5.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 31.25%
Indels: 0
Gaps: 0
DB: 2

US-09-854-133-587 (1-16) x US-08-858-311-6 (1-258)

QY 7 IleAspPheillelle 11
Db 172 ATAGATTCATAATT 186

RESULT 41
US-09-438-906-15/C
; Sequence 15, Application US/09438906
; Patent No. 6465185
; GENERAL INFORMATION:
; APPLICANT: Goldfine, Ira
; APPLICANT: Trischitta, Vincenzo
; APPLICANT: Pizzuti, Antonio
; APPLICANT: Vigneri, Riccardo
; APPLICANT: Frittitta, Lucia
; TITLE OF INVENTION: Polymorphic Human PC-1 Sequences
; FILE REFERENCE: 9076-089
; CURRENT APPLICATION NUMBER: US/09/438,906
; CURRENT FILING DATE: 1999-11-18
```

; PRIOR APPLICATION NUMBER: 60/108,853  
 ; PRIOR FILING DATE: 1998-11-18  
 ; NUMBER OF SEQ ID NOS: 35  
 ; SOFTWARE: EastSeq for Windows Version 4.0  
 ; SEQ ID NO 15  
 ; LENGTH: 313  
 ; TYPE: DNA  
 ; ORGANISM: H. sapiens  
 US-09-438-906-15

## Alignment Scores:

Pred. No.:	311	Length:	313
Score:	5.00	Matches:	5
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	31.25%	Indels:	0
DB:	4	Gaps:	0

US-09-854-133-587 (1-16) x US-09-438-906-15 (1-313)

QY 8 AspPheilleillephe 12  
 |||||  
 Db 215 GATTGTTTATTTT 201

## RESULT 42

US-09-134-001C-2354/c  
 ; Sequence 2354, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: GTC-007  
 ; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US/09/134,001C  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 2354  
 ; LENGTH: 390  
 ; TYPE: DNA  
 ; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-2354

## Alignment Scores:

Pred. No.:	376	Length:	390
Score:	5.00	Matches:	5
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	31.25%	Indels:	0
DB:	4	Gaps:	0

US-09-854-133-587 (1-16) x US-09-134-001C-2354 (1-390)

QY 7 IleAspPheilleille 11  
 |||||  
 Db 146 ATTGACTTCATCATA 132

## RESULT 43

US-09-134-001C-2395  
 ; Sequence 2395, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: GTC-007

; PRIOR APPLICATION NUMBER: US/09/134,001C  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 2395  
 ; LENGTH: 405  
 ; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-2395

## Alignment Scores:

Pred. No.:	389	Length:	405
Score:	5.00	Matches:	5
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	31.25%	Indels:	0
DB:	4	Gaps:	0

US-09-854-133-587 (1-16) x US-09-134-001C-2395 (1-405)

QY 7 IleAspPheilleille 11  
 |||||  
 Db 265 ATTGACTTCATCATA 279

## RESULT 44

US-08-469-667-21/c  
 ; Sequence 21, Application US/08469667  
 ; Patent No. 5733748  
 ; GENERAL INFORMATION:

; APPLICANT: Yu, Guo-Liang  
 ; APPLICANT: Rosen, Craig  
 ; TITLE OF INVENTION: Colon Specific Genes and Proteins  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
 ; ADDRESSEE: Stewart & Olstein  
 ; STREET: 6 Becker Farm Road  
 ; CITY: Roseland  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07068-1739

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/469,667  
 ; FILING DATE: 06-JUN-1995

## CLASSIFICATION:

; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ferraro, Gregory D.  
 ; REGISTRATION NUMBER: 36,134

## REFERENCE/DOCKET NUMBER:

; REFERENCE/DOCKET NUMBER: 325800-435

## TELEPHONE:

; TELEPHONE: 201-994-1700

## TELEFAX:

; TELEFAX: 201-994-1744

## INFORMATION FOR SEQ ID NO:

; INFORMATION FOR SEQ ID NO: 21:

## SEQUENCE CHARACTERISTICS:

; LENGTH: 409 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-469-667-21

## Alignment Scores:

Pred. No.:	392	Length:	409
Score:	5.00	Matches:	5
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	31.25%	Indels:	0
DB:	1	Gaps:	0

us-09-854-133-587.oligo.rni

Tue May 13 12:12:58 2003

US-09-854-133-587 (1-16) x US-08-469-667-21 (1-409)

QY 11 IlePheTrpIlePhe 15  
 |||||  
 Db 171 ATTTCGGATATT 157

## RESULT 45

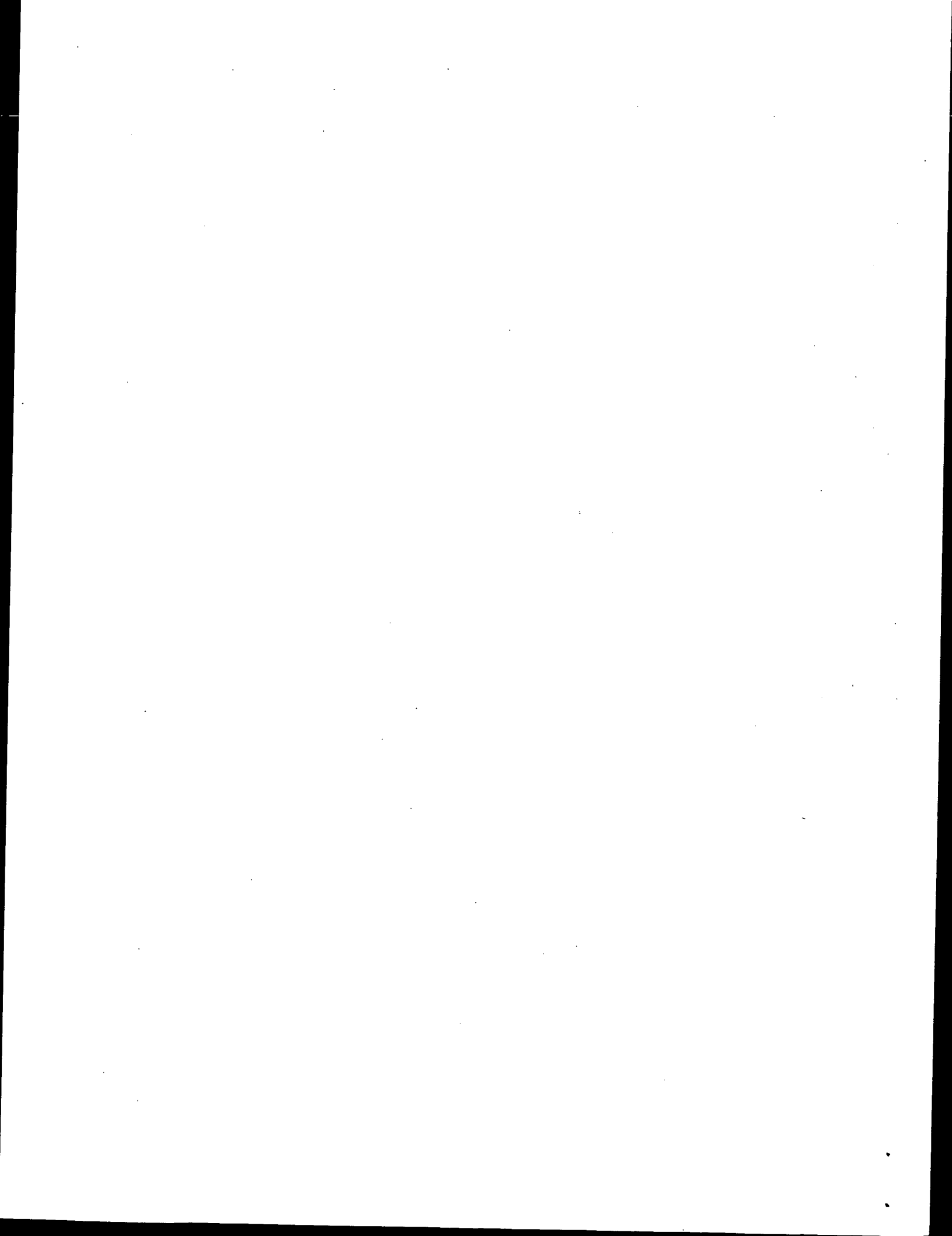
US-09-224-110-21/c  
 ; Sequence 21, Application US/09224110  
 ; Patent No. 6337195  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yu, Guo-Liang  
 ; APPLICANT: Rosen, Craig  
 ; TITLE OF INVENTION: Colon Specific Genes and Proteins  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
 ; ADDRESSEE: Stewart & Olstein  
 ; STREET: 6 Becker Farm Road  
 ; CITY: Roseland  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07068-1739  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/224,110  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/469,667  
 ; FILING DATE: 06-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ferraro, Gregory D.  
 ; REGISTRATION NUMBER: 36,134  
 ; REFERENCE/DOCKET NUMBER: 325800-435  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201-994-1700  
 ; TELEFAX: 201-994-1744  
 ; INFORMATION FOR SEQ ID NO: 21:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 409 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: CDNA  
 ; US-09-224-110-21

Alignment Scores:	392	Length:	409
Pred. No.:	5.00	Matches:	5
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	31.25%	Gaps:	0
DB:	4		

US-09-854-133-587 (1-16) x US-09-224-110-21 (1-409)

QY 11 IlePheTrpIlePhe 15  
 |||||  
 Db 171 ATTTCGGATATT 157

Search completed: May 11, 2003, 17:17:52  
 Job time : 18.9204 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 17:18:00 ; Search time 18.9735 Seconds  
(without alignments)  
1047.953 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANGIDFIIFWIFW 16

Scoring table:  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Egapop 6.0 , Egapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 783854 seqs, 621352466 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 1563015

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters: -DEV-rlh  
-MODEL-framer-p2n.model -US09854133/runat\_05052003\_174135\_809/app\_query.fasta\_1.462  
-DB-Published Applications NA -QPM-fastap -SUFFIX-oligo.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX-oligo  
-TRANS-human40.cdi -LIST=50 -DOCALIGN=200 -THR SCORE-quality -THR MIN=1  
-ALIGN=45 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USPR=US09854133 -CGN1\_1\_117 -runat\_05052003\_174135\_809  
-NCPU=6 -ICPU=3 -NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications, NA: \*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq: \*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq: \*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq: \*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq: \*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq: \*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq: \*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq: \*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq: \*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq: \*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq: \*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq: \*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq: \*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq: \*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	100.0	337	9	US-09-854-133-442
2	16	100.0	337	10	US-09-738-973-442
3	16	100.0	2239	9	US-09-854-133-440
4	16	100.0	2239	10	US-09-738-973-440

## RESULT 1

US-09-854-133-442  
; Sequence 442, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10

## ALIGNMENTS

5	16	100.0	5981	9	US-09-854-133-441	Sequence 441, App
6	16	100.0	5981	10	US-09-738-973-441	Sequence 441, App
c	7	43.8	170834	10	US-09-835-232-7	Sequence 7, Appli
8	6	37.5	26	10	US-09-006-298-7	Sequence 7, Appli
c	9	37.5	250	10	US-09-867-701-51	Sequence 51, Appli
10	6	37.5	277	9	US-10-076-622-101	Sequence 101, App
11	6	37.5	277	10	US-09-604-287A-101	Sequence 101, App
12	6	37.5	277	10	US-09-339-338-101	Sequence 101, App
13	6	37.5	277	12	US-10-007-805-101	Sequence 101, App
14	6	37.5	283	10	US-09-867-701-10058	Sequence 10058, A
15	6	37.5	390	9	US-09-796-692-9005	Sequence 9005, Ap
16	6	37.5	390	9	US-10-040-862-9005	Sequence 9005, Ap
c	17	37.5	416	9	US-09-918-995-5321	Sequence 5321, Ap
c	18	37.5	462	9	US-09-918-995-13913	Sequence 13913, A
19	6	37.5	498	10	US-09-880-107-1080	Sequence 1080, Ap
20	6	37.5	505	10	US-09-998-598-1101	Sequence 1101, Ap
21	6	37.5	579	10	US-09-764-877-243	Sequence 243, App
22	6	37.5	600	10	US-09-864-761-8345	Sequence 8345, Ap
23	6	37.5	696	10	US-09-974-300-6399	Sequence 6399, Ap
24	6	37.5	989	10	US-09-867-550-1741	Sequence 1741, Ap
25	6	37.5	1188	10	US-09-741-669-167	Sequence 167, App
26	6	37.5	1260	10	US-09-815-242-4288	Sequence 4288, Ap
27	6	37.5	1263	10	US-09-815-242-8550	Sequence 8550, Ap
c	28	37.5	1263	10	US-09-815-242-8719	Sequence 8719, Ap
29	6	37.5	1268	10	US-09-070-927A-557	Sequence 557, App
30	6	37.5	1330	9	US-09-746-660A-11	Sequence 11, Appli
31	6	37.5	1455	9	US-10-142-231-43	Sequence 43, Appli
32	6	37.5	1455	9	US-10-142-231-44	Sequence 44, Appli
33	6	37.5	1491	9	US-09-746-660A-9	Sequence 9, Appli
34	6	37.5	1524	9	US-10-142-231-52	Sequence 52, Appli
35	6	37.5	1623	10	US-09-974-300-6367	Sequence 6367, Ap
36	6	37.5	1786	10	US-09-764-877-4007	Sequence 4007, Ap
37	6	37.5	2000	9	US-09-938-842A-3762	Sequence 3762, Ap
38	6	37.5	2000	9	US-09-938-842A-3983	Sequence 3983, Ap
39	6	37.5	2000	9	US-09-938-842A-4384	Sequence 4384, Ap
c	40	37.5	2524	7	US-08-781-986A-360	Sequence 360, App
c	41	37.5	2692	9	US-10-092-154-1254	Sequence 1254, Ap
c	42	37.5	2692	10	US-09-764-847-1254	Sequence 1254, Ap
c	43	37.5	2694	9	US-10-092-154-1255	Sequence 1255, Ap
c	44	37.5	2694	10	US-09-764-847-1255	Sequence 1255, Ap
45	6	37.5	2912	9	US-10-071-766-61	Sequence 61, Appli
46	6	37.5	3720	9	US-10-076-622-576	Sequence 576, App
47	6	37.5	3720	12	US-10-007-805-576	Sequence 576, App
48	6	37.5	5285	9	US-10-239-676-157	Sequence 157, App
49	6	37.5	5508	10	US-09-954-456-943	Sequence 943, App
50	6	37.5	11230	9	US-10-150-821-3	Sequence 3, Appli
c	51	37.5	11230	10	US-09-911-842-3	Sequence 3, Appli
c	52	37.5	22786	10	US-09-764-877-3727	Sequence 3727, Ap
53	6	37.5	27332	9	US-10-092-154-1555	Sequence 1555, Ap
54	6	37.5	27332	10	US-09-764-847-1555	Sequence 1555, Ap
55	6	37.5	55795	10	US-09-880-107-1543	Sequence 1543, Ap
56	6	37.5	124884	9	US-09-913-514-1	Sequence 1, Appli
57	6	37.5	125157	9	US-09-913-514-2	Sequence 2, Appli
58	6	37.5	640681	10	US-09-790-988-1	Sequence 1, Appli
c	59	31.2	17	9	US-09-818-875-1906	Sequence 1906, Ap
60	5	31.2	17	9	US-09-818-875-1907	Sequence 1907, Ap

; CURRENT APPLICATION NUMBER: US/09/854,133  
 ; CURRENT FILING DATE: 2001-05-11  
 ; NUMBER OF SEQ ID NOS: 735  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 442  
 ; LENGTH: 337  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-854-133-442

Alignment Scores:  
 Pred. No.: 111e-09 Length: 337  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-854-133-442 (1-337)

QY 1 PheGlnAlaAsnCysGlyleAspPheIlePheTrpIlePheTrp 16  
 Db 107 TTCCAGGCCAATTGGCATAGATTTATCATATCTGGATTTTGG 154

## RESULT 2

US-09-738-973-442  
 ; Sequence 442, Application US/09738973  
 ; Patent No. US20020110563A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indrias, Carol Yoseph  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Elliot, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14  
 ; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 442  
 ; LENGTH: 337  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-738-973-442

Alignment Scores:  
 Pred. No.: 111e-09 Length: 337  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-738-973-442 (1-337)

QY 1 PheGlnAlaAsnCysGlyleAspPheIlePheTrpIlePheTrp 16  
 Db 107 TTCCAGGCCAATTGGCATAGATTTATCATATCTGGATTTTGG 154

## RESULT 3

US-09-854-133-440  
 ; Sequence 440, Application US/09854133  
 ; Publication No. US20020183499A1  
 ; GENERAL INFORMATION:

; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Secrist, Heather  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
 ; FILE REFERENCE: 210121.475C10  
 ; CURRENT APPLICATION NUMBER: US/09/854,133  
 ; CURRENT FILING DATE: 2001-05-11  
 ; NUMBER OF SEQ ID NOS: 735  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 440  
 ; LENGTH: 2239  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-854-133-440

Alignment Scores:  
 Pred. No.: 7.04e-09 Length: 2239  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-854-133-440 (1-2239)

QY 1 PheGlnAlaAsnCysGlyleAspPheIlePheTrpIlePheTrp 16  
 Db 104 TTCCAGGCCAATTGGCATAGATTTATCATATCTGGATTTTGG 151

## RESULT 4

US-09-738-973-440  
 ; Sequence 440, Application US/09738973  
 ; Patent No. US20020110563A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indrias, Carol Yoseph  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Elliot, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14  
 ; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 440  
 ; LENGTH: 2239  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-738-973-440

Alignment Scores:  
 Pred. No.: 7.04e-09 Length: 2239  
 Score: 16.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-738-973-440 (1-2239)

QY 1 PheGlnAlaAsnCysGlyleAspPheIlePheTrpIlePheTrp 16



Db 104 TTCAGGCCAATTGGCAGATTTATCATATCTGGATTTTGG 151  
|||||

## RESULT 5

US-09-854-133-441  
; Sequence 441, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 441  
; LENGTH: 5981  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-854-133-441

Alignment Scores:  
Pred. No.: 1,84e-08 Length: 5981  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-854-133-441 (1-5981)

Qy 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrp 16  
|||||

Db 102 TTCAGGCCAATTGGCAGATTTATCATATCTGGATTTTGG 149  
|||||

## RESULT 6

US-09-738-973-441  
; Sequence 441, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Manion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 441  
; LENGTH: 5981  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-738-973-441

Alignment Scores:  
Pred. No.: 1,84e-08 Length: 5981  
Score: 16.00 Matches: 16

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-738-973-441 (1-5981)

Qy 1 PheGlnAlaAsnCysGlyIleAspPheIlePheTrp 16  
|||||

Db 102 TTCAGGCCAATTGGCAGATTTATCATATCTGGATTTTGG 149  
|||||

## RESULT 7

US-09-835-232-7/c  
; Sequence 7, Application US/09835232  
; Patent No. US20020098489A1  
; GENERAL INFORMATION:  
; APPLICANT: Leder, Philip  
; APPLICANT: Leader, Benjamin  
; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 00383/052002  
; CURRENT APPLICATION NUMBER: US/09/835,232  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: US 60/196,811  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 170834  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(170834)  
; OTHER INFORMATION: n= A,T,C, or G  
US-09-835-232-7

Alignment Scores:  
Pred. No.: 3,32e+03 Length: 170834  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-835-232-7 (1-170834)

Qy 10 llellePheTrpIlePheTrp 16  
|||||

Db 95470 ATCATCTTTTGGATTTTCTGG 95450  
|||||

## RESULT 8

US-09-006-298-7  
; Sequence 7, Application US/09006298  
; Patent No. US2002008224A1  
; GENERAL INFORMATION:  
; APPLICANT: Jolly, Douglas J.  
; APPLICANT: Moore, Margaret D.  
; APPLICANT: Chada, Sunil  
; TITLE OF INVENTION: NON-IMMUNOGENIC PRODRUGS AND SELECTABLE  
; TITLE OF INVENTION: MARKERS FOR USE IN GENE THERAPY  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/006,298
; FILING DATE: 13-JAN-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMahsters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 930049.459
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-006-298-7

```

```

Alignment Scores:
Pred. No.: 7.68 Length: 26
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

```

US-09-854-133-587 (1-16) x US-09-006-298-7 (1-26)

```

QY 6 Glytleasphellelle 11
Db 4 GGGATCGATTTCATC 21

```

## RESULT 9

```

US-09-867-701-51/c
; Sequence 51, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-51

```

```

Alignment Scores:
Pred. No.: 70.1 Length: 250
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

```

US-09-854-133-587 (1-16) x US-09-867-701-51 (1-250)

```

QY 10 IlellePheTrpIlePhe 15
Db 100 ATAATTTCTGGATTTT 83

```

## RESULT 10

```

US-10-076-622-101
; Sequence 101, Application US/10076622
; Publication No. US20030023036A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-076-622-101

```

```

Alignment Scores:
Pred. No.: 77.5 Length: 277
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0

```

US-09-854-133-587 (1-16) x US-10-076-622-101 (1-277)

```

QY 10 IlellePheTrpIlePhe 15
Db 96 ATTATATTGGATCTTC 113

```

## RESULT 11

```

US-09-604-287A-101
; Sequence 101, Application US/09604287A
; Patent No. US20020064872A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-604-287A-101

```

```

Alignment Scores:
Pred. No.: 77.5 Length: 277
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

```

US-09-854-133-587 (1-16) x US-09-604-287A-101 (1-277)

```

QY 10 IlellePheTrpIlePhe 15
Db 96 ATTATATTGGATCTTC 113

```

## RESULT 12

```

US-09-339-338-101
; Sequence 101, Application US/09339338A
; Patent No. US20020102602A1

```

Tue May 13 12:12:59 2003

us-09-854-133-587.oligo.rnpb

```

; GENERAL INFORMATION:
; APPLICANT: Yuqiu, Jiang
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.470C2
; CURRENT APPLICATION NUMBER: US/09/339,338A
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-339-338-101

Alignment Scores:
Pred. No.: 77.5 Length: 277
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-339-338-101 (1-277)
QY 10 lilelPheTrpIlePhe 15
DB 96 ATTATATTGGATCTTC 113

RESULT 13
US-10-007-805-101
; Sequence 101, Application US/10007805
; Patent No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepner, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-007-805-101

Alignment Scores:
Pred. No.: 77.5 Length: 277
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x US-10-007-805-101 (1-277)
QY 10 lilelPheTrpIlePhe 15
DB 96 ATTATATTGGATCTTC 113
```

```

DB 96 ATTATATTGGATCTTC 113

RESULT 14
US-09-867-701-10058
; Sequence 10058, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10058
; LENGTH: 283
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-701-10058

Alignment Scores:
Pred. No.: 79.1 Length: 283
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-867-701-10058 (1-283)
QY 9 PheIlelPheTrpIle 14
DB 21 TTTATATTATTGGATA 38

RESULT 15
US-09-796-692-9005
; Sequence 9005, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
```

```
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9005
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-9005
```

```
Alignment Scores:
Pred. No.: 108
Score: 6.00
Length: 390
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 37.50%
DB: 9
Matches: 6
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-854-133-587 (1-16) x US-09-796-692-9005 (1-390)
```

```
QY 10 IleilePheTrillePhe 15
|||||
Db 229 ATTATATTGGATCTTC 246
```

## RESULT 16

```
; Sequence 9005, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; PRIOR FILING DATE: 2001-11-06
; PRIOR FILING DATE: 2000-03-01
; PRIOR FILING DATE: 2000-03-01
; PRIOR FILING DATE: 2000-03-17
; PRIOR FILING DATE: 2000-04-27
; PRIOR FILING DATE: 2000-04-27
; PRIOR FILING DATE: 2000-04-28
; PRIOR FILING DATE: 2000-04-28
; PRIOR FILING DATE: 2000-04-28
; PRIOR FILING DATE: 2000-05-01
; PRIOR FILING DATE: 2000-05-01
; PRIOR FILING DATE: 2000-05-04
; PRIOR FILING DATE: 2000-05-04
; PRIOR FILING DATE: 2000-05-22
; PRIOR FILING DATE: 2000-07-14
; PRIOR FILING DATE: 2000-08-03
; PRIOR FILING DATE: 2000-08-03
; PRIOR FILING DATE: 2000-08-04
; PRIOR FILING DATE: 2000-08-07
; PRIOR FILING DATE: 2000-08-07
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9005
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-9005
```

## Alignment Scores:

```
Pred. No.: 108
Score: 6.00
Length: 390
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 37.50%
DB: 9
Matches: 6
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
```

```
US-09-854-133-587 (1-16) x US-10-040-862-9005 (1-390)
```

```
QY 10 IleilePheTrillePhe 15
|||||
Db 229 ATTATATTGGATCTTC 246
```

## RESULT 17

```
; Sequence 9321, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5321
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (1)-(416)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-5321
```

## Alignment Scores:

```
Pred. No.: 115
Score: 6.00
Length: 416
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 37.50%
DB: 9
Matches: 6
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
```

```
US-09-854-133-587 (1-16) x US-09-918-995-5321 (1-416)
```

```
QY 6 GlyVileAspPheillelle 11
|||||
Db 105 GGCATTGACITCATTATA 88
```

## RESULT 18

```
; Sequence 13913, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13913
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
```

; LOCATION: (1)...(462)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-13913

Alignment Scores: Length: 462  
Pred. No.: 128 Matches: 6  
Score: 6.00 Conservatives: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB:

US-09-854-133-587 (1-16) x US-09-918-995-13913 (1-462)

QY 10 ilellePheTrillePhe 15

DB 351 ATTATTCTGGATATTC 334

RESULT 19

US-09-880-107-1080  
; Sequence 1080, Application US/09880107  
; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherf, Uwe

; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880,107

; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: US 60/237,054

; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1080

; LENGTH: 498

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA443756

US-09-880-107-1080

Alignment Scores: Length: 498  
Pred. No.: 137 Matches: 6  
Score: 6.00 Conservatives: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB:

US-09-854-133-587 (1-16) x US-09-880-107-1080 (1-498)

QY 6 GlyIleAspPheIleIle 11

DB 48 GGCATTGACTTCATTATA 65

RESULT 20

US-09-998-598-1101

; Sequence 1101, Application US/09998598

; Patent No. US20020150922A1

; GENERAL INFORMATION:

; APPLICANT: Stolk, John A.

; APPLICANT: Xu, Jiangchun

; APPLICANT: Chenault, Ruth A.

; APPLICANT: Meagher, Madelein Joy

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.561

; CURRENT APPLICATION NUMBER: US/09/998,598

; CURRENT FILING DATE: 2001-11-16

; NUMBER OF SEQ ID NOS: 2606  
; SOFTWARE: Corixa Invention Disclosure Database  
; SEQ ID NO 1101  
; LENGTH: 505  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-998-598-1101

Alignment Scores: Length: 505  
Pred. No.: 139 Matches: 6  
Score: 6.00 Conservatives: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB:

US-09-854-133-587 (1-16) x US-09-998-598-1101 (1-505)

QY 7 ileAspPheIleIlePhe 12

DB 369 ATTGATTTTATTATTTT 386

RESULT 21

US-09-764-877-243

; Sequence 243, Application US/09764877

; Patent No. US20020147140A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC005

; CURRENT APPLICATION NUMBER: US/09/764,877

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 4031

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 243

; LENGTH: 579

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (474)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (579)

; OTHER INFORMATION: n equals a,t,g, or c

US-09-764-877-243

Alignment Scores: Length: 579  
Pred. No.: 159 Matches: 6  
Score: 6.00 Conservatives: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB:

US-09-854-133-587 (1-16) x US-09-764-877-243 (1-579)

QY 1 PheGlnAlaAsnCysGly 6

DB 344 TTCCAAGCAACTGGCGT 361

RESULT 22

US-09-864-761-8345

; Sequence 8345, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USED

; CURRENT APPLICATION NUMBER: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aeomica-x-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 8345  
LENGTH: 600  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC015971.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.8  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.8  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.5  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 8.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.5  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.7  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6

Alignment Scores:  
Pred. No.: 165  
Score: 6.00  
Length: 600  
Matches: 6  
Conservative: 0  
Best Local Similarity: 100.00%  
Query Match: 37.50%  
Indels: 0  
Gaps: 0  
DB: 10

US-09-854-133-587 (1-16) x US-09-864-761-8345 (1-600)

Qy 6 GlyleAspPheIlelle 11  
Db 489 GGCATTGACTTCATCATT 506

RESULT 23

US-09-974-300-6399/c

Sequence 6399, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
APPLICANT: Clausen, Ib Groth  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
TITLE OF INVENTION: Expression  
FILE REFERENCE: 10085.500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680,598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279,526  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6399  
LENGTH: 696  
TYPE: DNA  
ORGANISM: Bacillus clausii  
US-09-974-300-6399

Alignment Scores:  
Pred. No.: 191  
Score: 6.00  
Length: 696  
Matches: 6  
Conservative: 0  
Best Local Similarity: 100.00%  
Query Match: 37.50%  
Indels: 0  
Gaps: 0  
DB: 10

US-09-854-133-587 (1-16) x US-09-974-300-6399 (1-696)

Qy 10 IleIlePheTTPlePhe 15  
Db 587 ATAATTTCCTGGATTTT 570

RESULT 24

US-09-867-550-1741  
Sequence 1741, Application US/09867550  
Patent No. US20020082206A1  
GENERAL INFORMATION:

APPLICANT: Leach, Martin D.  
APPLICANT: Mehrahan, Fuad,  
APPLICANT: Conley, Pamela  
APPLICANT: Law, Debbie  
APPLICANT: Topper, James  
TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells  
FILE REFERENCE: 21402-013 (Cura-313)  
CURRENT APPLICATION NUMBER: US/09/867,550  
CURRENT FILING DATE: 2001-09-20  
PRIOR APPLICATION NUMBER: USSN 60/208,427  
PRIOR FILING DATE: 2000-05-30  
NUMBER OF SEQ ID NOS: 2125  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1741  
LENGTH: 989  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)  
OTHER INFORMATION: Wherein n is one of a or t or c or g

Alignment Scores:

Pred. No.: 269  
Score: 6.00  
Length: 989  
Matches: 6  
Conservative: 0  
Best Local Similarity: 100.00%  
Query Match: 37.50%  
Indels: 0  
Gaps: 0  
DB: 10

US-09-867-550-1741

## us-09-854-133-587.oligo.rnpb

Tue May 13 12:12:59 2003

```

US-09-854-133-587 (1-16) x US-09-867-550-1741 (1-989)
QY 11 IlePheTrpIlePheTrp 16
Db 557 ATATTGGATCTCTGG 574

RESULT 25
US-09-741-669-167/c
; Sequence 167, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1188)
US-09-741-669-167
Alignment Scores: 321 Length: 1188
Pred. No.: 6 Matches: 6
Score: 6.00 Conservativeness: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 37.50% Gaps: 0
DB: 10

US-09-854-133-587 (1-16) x US-09-741-669-167 (1-1188)
QY 6 GlyIleAspPheIleIle 11
Db 782 GGTATAGATTTTATTAAT 765

RESULT 26
US-09-815-242-4288/c
; Sequence 4288, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8550
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1263)
US-09-815-242-8550
Alignment Scores: 341 Length: 1263
Pred. No.: 6 Matches: 6
Score: 6.00

```

Percent Similarity: 100.00%    Conservative: 0  
 Best Local Similarity: 100.00%    Mismatches: 0  
 Query Match: 37.50%    Indels: 0  
 DB: 10    Gaps: 0

US-09-854-133-587 (1-16) x US-09-815-242-8550 (1-1263)

QY 7 IleAspPheIleIlePhe 12

Db 1191 ATCGATTATTATTTTC 1174

RESULT 28

US-09-815-242-8719/c  
 ; Sequence 8719, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Karl L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 8719  
 ; LENGTH: 1263  
 ; TYPE: DNA  
 ; ORGANISM: Staphylococcus aureus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(1263)

US-09-815-242-8719

Alignment Scores:  
 Pred. No.: 341    Length: 1263  
 Score: 6.00    Matches: 6  
 Percent Similarity: 100.00%    Conservative: 0  
 Best Local Similarity: 100.00%    Mismatches: 0  
 Query Match: 37.50%    Indels: 0  
 DB: 10    Gaps: 0

US-09-854-133-587 (1-16) x US-09-815-242-8719 (1-1263)

QY 7 IleAspPheIleIlePhe 12

Db 1191 ATCGATTATTATTTTC 1174

RESULT 29

US-09-070-927A-557  
 ; Sequence 557, Application US/09070927A  
 ; Patent No. US20020120116A1  
 ; GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch  
 Patrick J. Dillon  
 Steven Barash  
 TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypepti  
 NUMBER OF SEQUENCES: 982  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/070,927A  
 FILING DATE: 04-May-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/046,655  
 FILING DATE: 1997-05-16  
 APPLICATION NUMBER: 60/044,031  
 FILING DATE: 1997-05-06  
 APPLICATION NUMBER: 60/066,009  
 FILING DATE: 1997-11-14  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kenley K. Hoover  
 REGISTRATION NUMBER: 40,302  
 REFERENCE/DOCKET NUMBER: PB369  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 TELEFAX: (301) 309-8512  
 INFORMATION FOR SEQ ID NO: 557:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1268 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 557:  
 US-09-070-927A-557  
 Alignment Scores:  
 Pred. No.: 342    Length: 1268  
 Score: 6.00    Matches: 6  
 Percent Similarity: 100.00%    Conservative: 0  
 Best Local Similarity: 100.00%    Mismatches: 0  
 Query Match: 37.50%    Indels: 0  
 DB: 10    Gaps: 0  
 US-09-854-133-587 (1-16) x US-09-070-927A-557 (1-1268)  
 QY 6 GlyIleAspPheIleIle 11  
 Db 704 GGAATAGACTTCATTATC 721  
 RESULT 30  
 US-09-746-660A-11  
 ; Sequence 11, Application US/09746660A  
 ; Publication No. US20030049804A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pompejus, Markus  
 ; APPLICANT: Kroger, Burkhard  
 ; APPLICANT: Schroder, Hartwig  
 ; APPLICANT: Zelder, Oskar  
 ; APPLICANT: Haberhauer, Gregor  
 ; APPLICANT: Kim, Jun-Won  
 ; APPLICANT: Lee, Heung-Schick  
 ; APPLICANT: Hwang, Byung-Joon  
 ; TITLE OF INVENTION: CORINEBACTERIUM GLUTAMICUM GENES ENCODING  
 ; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS



Tue May 13 12:12:59 2003

us-09-854-133-587.oligo.rnpb

FILE REFERENCE: BGI-121CP2  
CURRENT APPLICATION NUMBER: US/09/746,660A  
CURRENT FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 09/606740  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 09/603124  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 60/141031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: 60/142101  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: 60/148613  
PRIOR FILING DATE: 1999-08-12  
PRIOR APPLICATION NUMBER: 60/187970  
PRIOR FILING DATE: 2000-03-09  
PRIOR APPLICATION NUMBER: DE 19931420.9  
PRIOR FILING DATE: 1999-07-08  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: PatentIn Vers. 2.0  
SEQ ID NO 11  
LENGTH: 1330  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (101)..(1330)  
OTHER INFORMATION: FRXA01009  
US-09-746-660A-11

Alignment Scores:  
Pred. No.: 359 Length: 1330  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-746-660A-11 (1-1330)

QY 8 AspPheIleIlePheTrp 13  
|||||  
DB 70 GACTTCATATATTGG 87

RESULT 31  
US-10-142-231-43  
Sequence 43, Application US/10142231  
Publication No. US2003007796A1  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney et al.  
TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES  
FILE REFERENCE: 62773  
CURRENT APPLICATION NUMBER: US/10/142,231  
CURRENT FILING DATE: 2002-05-08  
PRIOR APPLICATION NUMBER: 60/165,250  
PRIOR FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Vers. 2.1  
SEQ ID NO 43  
LENGTH: 1455  
TYPE: DNA  
ORGANISM: Taxus cuspidata  
US-10-142-231-43  
Alignment Scores:  
Pred. No.: 391 Length: 1455  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-142-231-43 (1-1455)

QY 8 AspPheIleIlePheTrp 13  
|||||  
DB 602 GACTTCATCATCTTTGG 619  
RESULT 32  
US-10-142-231-44  
Sequence 44, Application US/10142231  
Publication No. US2003007796A1  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney et al.  
TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES  
FILE REFERENCE: 62773  
CURRENT APPLICATION NUMBER: US/10/142,231  
CURRENT FILING DATE: 2002-05-08  
PRIOR APPLICATION NUMBER: 60/165,250  
PRIOR FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Vers. 2.1  
SEQ ID NO 44  
LENGTH: 1455  
TYPE: DNA  
ORGANISM: Taxus cuspidata  
US-10-142-231-44

Alignment Scores:  
Pred. No.: 391 Length: 1455  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-142-231-44 (1-1455)

QY 8 AspPheIleIlePheTrp 13  
|||||  
DB 602 GACTTCATCATCTTTGG 619

RESULT 33  
US-09-746-660A-9  
Sequence 9, Application US/09746660A  
Publication No. US2003004980A1  
GENERAL INFORMATION:  
APPLICANT: Pompejus, Markus  
APPLICANT: Kroger, Burkhard  
APPLICANT: Schroder, Hartwig  
APPLICANT: Zelder, Oskar  
APPLICANT: Haberhauer, Gregor  
APPLICANT: Kim, Jun-Won  
APPLICANT: Lee, Heung-Schick  
APPLICANT: Hwang, Byung-Joon  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
METABOLIC PATHWAY PROTEINS  
FILE REFERENCE: BGI-121CP2  
CURRENT APPLICATION NUMBER: US/09/746,660A  
CURRENT FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 09/606740  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 09/603124  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 60/141031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: 60/142101  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: 60/148613  
PRIOR FILING DATE: 1999-08-12  
PRIOR APPLICATION NUMBER: 60/187970  
PRIOR FILING DATE: 2000-03-09  
PRIOR APPLICATION NUMBER: DE 19931420.9  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: PatentIn Vers. 2.0  
SEQ ID NO 9

```

; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1468)
; OTHER INFORMATION: RXS02970
US-09-746-660A-9

Alignment Scores:
Pred. No.: 401          Length: 1491
Score: 6.00           Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-746-660A-9 (1-1491)
QY 8 AspPheIleIlePheTrp 13
Db 70 GACTTCATCATATTTTGG 87

RESULT 34
US-10-142-231-52
; Sequence 52, Application US/10142231
; Publication No. US2003007796A1
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
; FILE REFERENCE: 82773
; CURRENT APPLICATION NUMBER: US/10/142,231
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 60/165,250
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Taxus cuspidata
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(1524)
; OTHER INFORMATION: "n" equals any oligonucleotide
US-10-142-231-52

Alignment Scores:
Pred. No.: 410          Length: 1524
Score: 6.00           Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-142-231-52 (1-1524)
QY 8 AspPheIleIlePheTrp 13
Db 614 GACTTCATCATCTTTTGG 631

RESULT 35
US-09-974-300-6367/c
; Sequence 6367, Application US/09974300
; Patent No. US2002014672A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene Expression
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05

```

```

; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6367
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6367

Alignment Scores:
Pred. No.: 436          Length: 1623
Score: 6.00           Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-974-300-6367 (1-1623)
QY 10 IleIlePheTrpIlePhe 15
Db 1011 ATAATTTTCGTGATTTT 994

RESULT 36
US-09-764-877-4007/c
; Sequence 4007, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4007
; LENGTH: 1786
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-4007

Alignment Scores:
Pred. No.: 478          Length: 1786
Score: 6.00           Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-764-877-4007 (1-1786)
QY 10 IleIlePheTrpIlePhe 15
Db 1460 ATAATATTTGGATTTT 1443

RESULT 37
US-09-938-842A-3762/c
; Sequence 3762, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING STRESS-REGULATED GENES OF PLANTS, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24

```

; PRIOR APPLICATION NUMBER: US 60/227,866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264,647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 3762  
 ; LENGTH: 2000  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 US-09-938-842A-3762

Alignment Scores: 534 Length: 2000  
 Pred. No.: 6.00 Matches: 6  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 37.50% Gaps: 0  
 DB: 9

US-09-854-133-587 (1-16) x US-09-938-842A-3762 (1-2000)

QY 10 lilellePheTrpIlePhe 15  
 Db 1337 ATTATATTTGGATATT 1320

RESULT 38

US-09-938-842A-3983  
 ; Sequence 3983, Application US/09938842A  
 ; Patent No. US20020160378A1

; GENERAL INFORMATION:  
 ; APPLICANT: Harper, Jeff  
 ; APPLICANT: Kreps, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 ; FILE REFERENCE: SCRIPI300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938,842A  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/227,866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264,647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 3983  
 ; LENGTH: 2000  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 US-09-938-842A-3983

Alignment Scores: 534 Length: 2000  
 Pred. No.: 6.00 Matches: 6  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 37.50% Gaps: 0  
 DB: 9

US-09-854-133-587 (1-16) x US-09-938-842A-3983 (1-2000)

QY 9 PheIlellePheTrpIle 14  
 Db 1609 TTTATATCTTTGGATT 1626

RESULT 39

US-09-938-842A-4984  
 ; Sequence 4984, Application US/09938842A  
 ; Patent No. US20020160378A1

; GENERAL INFORMATION:  
 ; APPLICANT: Harper, Jeff  
 ; APPLICANT: Kreps, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 ; FILE REFERENCE: SCRIPI300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938,842A  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/227,866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264,647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 3983  
 ; LENGTH: 2000  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 US-09-938-842A-3983

; APPLICANT: Harper, Jeff  
 ; APPLICANT: Kreps, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 ; FILE REFERENCE: SCRIPI300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938,842A  
 ; CURRENT FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/227,866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264,647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 4984  
 ; LENGTH: 2000  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 US-09-938-842A-4984

Alignment Scores: 534 Length: 2000  
 Pred. No.: 6.00 Matches: 6  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 37.50% Gaps: 0  
 DB: 9

US-09-854-133-587 (1-16) x US-09-938-842A-4984 (1-2000)

QY 10 lilellePheTrpIlePhe 15  
 Db 1546 ATTATATTTGGATATT 1563

RESULT 40

US-08-781-986A-360/c  
 ; Sequence 360, Application US/08781986A  
 ; Publication No. US20030054436A1

; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 5255  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/781,986A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Benson, Bob  
 ; REGISTRATION NUMBER: 30,446  
 ; REFERENCE/DOCKET NUMBER: PB248PP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 360:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2524 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-360
Alignment Scores:
Pred. No.: 670 Length: 2524
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 7 Gaps: 0

US-09-854-133-587 (1-16) x US-08-781-986A-360 (1-2524)
QY 7 IleasPhelellelle 12
| | | | | | | | | | | | | | | |
Db 1322 ATCGATTGACTTCATTATA 1305

RESULT 41
US-10-092-154-1254/c
; Sequence 1254, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1254
; LENGTH: 2692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1254
Alignment Scores:
Pred. No.: 714 Length: 2692
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-092-154-1254 (1-2692)
QY 6 GlyileAspPhelelle 11
| | | | | | | | | | | | | | | |
Db 2649 GGCATTGACTTCATTATA 2632

RESULT 42
US-09-764-847-1254/c
; Sequence 1254, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1254
; LENGTH: 2692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1254
Alignment Scores:
Pred. No.: 714 Length: 2692
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0

```

```

Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-764-847-1254 (1-2692)
QY 6 GlyileAspPhelelle 11
| | | | | | | | | | | | | | | |
Db 2649 GGCATTGACTTCATTATA 2632

RESULT 43
US-10-092-154-1255/c
; Sequence 1255, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1255
; LENGTH: 2694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1255
Alignment Scores:
Pred. No.: 715 Length: 2694
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-092-154-1255 (1-2694)
QY 6 GlyileAspPhelelle 11
| | | | | | | | | | | | | | | |
Db 2651 GGCATTGACTTCATTATA 2634

RESULT 44
US-09-764-847-1255/c
; Sequence 1255, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1255
; LENGTH: 2694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1255
Alignment Scores:
Pred. No.: 715 Length: 2694
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.50% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-764-847-1255 (1-2694)

```

us-09-854-133-587.oligo.rnpb

Tue May 13 12:12:59 2003

```

QY      6 Glyrlaspheillelle 11
      |||
Db 2651 GGCATTGACTTCATTATA 2634

RESULT 45
US-10-071-766-61
; Sequence 61, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
;   APPLICANT: Huel-Mei Chen
;   TITLE OF INVENTION: GENES EXPRESSED IN SENESCENCE
;   FILE REFERENCE: PA-0043 US
;   CURRENT APPLICATION NUMBER: US/10/071,766
;   CURRENT FILING DATE: 2002-02-07
;   NUMBER OF SEQ ID NOS: 144
;   SOFTWARE: PERL Program
;   SEQ ID NO 61
;   LENGTH: 2912
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020192678A1 037877.2
US-10-071-766-61

Alignment Scores:
Pred. No.:      771      Length:      2912
Score:          6.00      Matches:      6
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      37.50%      Indels: 0
DB:               9      Gaps: 0

US-09-854-133-587 (1-16) x US-10-071-766-61 (1-2912)

QY      10 IlellePheTrpIlePhe 15
      |||
Db 1275 ATTATATTTGGATCTTC 1292

Search completed: May 11, 2003, 18:46:09
Job time : 33.9734 secs

```



Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 16:29:05 ; Search time 242.973 seconds  
(without alignments)  
1066.486 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 32305347

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 60 summaries

Command line parameters:  
-MODEP=frame+p2n.model -DEV=xlh  
-O/cgn2\_1/USPQ.spool/US09854133/runat\_05052003\_174133\_718/app\_query.fasta\_1.462  
-DB-EST -QFMT=fastap SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=60  
-DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=plc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn\_1.1.2013@runat\_05052003\_174133\_718 -NCPU=6 -ICPU=3  
-NO\_XIPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FQAPOP=6 -FQAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :  
EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estom:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	8	50.0	876	12	BF693034 602080030
C 2	7	43.8	240	9	AU073495 AU073495
C 3	7	43.8	284	10	BB261269 BB261269
C 4	7	43.8	336	12	BE954331 UI-M-CEL-
C 5	7	43.8	438	12	BG446254 GA_EB003
C 6	7	43.8	440	12	BG446243 GA_EB003
C 7	7	43.8	450	17	AQ115066 CIT-HSP-2
C 8	7	43.8	452	9	AU270164 AU270164
C 9	7	43.8	457	9	AU269237 AU269237
C 10	7	43.8	460	12	BG543829 E1546 Chi
C 11	7	43.8	465	17	AZ246141 PM3-HT090
C 12	7	43.8	506	12	BF356741 EST461944
C 13	7	43.8	511	12	BG097509 BGS28250
C 14	7	43.8	529	13	BJ528250 BJ528250
C 15	7	43.8	572	12	BE971441 BE971441
C 16	7	43.8	598	12	BG571089 BGS71089
C 17	7	43.8	612	17	BH539329 BGS39329
C 18	7	43.8	617	17	AZ316177 IM0033A16
C 19	7	43.8	620	17	AZ937525 ZM0195C17
C 20	7	43.8	642	14	BU008061 QG06F17-Y
C 21	7	43.8	675	14	BQ868209 QGD13F14
C 22	7	43.8	710	9	AU136235 AU136235
C 23	7	43.8	717	17	BH695614 BHS2214TF
C 24	7	43.8	747	17	BH273281 CH230-23H
C 25	7	43.8	753	17	BH422412 BOHSD57TR
C 26	7	43.8	794	17	BH566426 BOHSG71TR
C 27	7	43.8	799	17	AQ898772 HS_3130_A
C 28	7	43.8	864	17	BH152622 ENT1W09TR
C 29	7	43.8	879	17	AZ692543 ENT1W09TR
C 30	7	43.8	882	17	AZ692069 ENTME42TR
C 31	7	43.8	883	12	BF675705 602083612
C 32	7	43.8	908	17	AZ545454 ENTG311TR
C 33	7	43.8	908	17	BH130713 ENTN171F
C 34	7	43.8	971	12	BF215221 601846206
C 35	7	43.8	996	10	BE250032 600943035
C 36	7	43.8	1077	13	B1260864 602971086
C 37	6	37.5	91	14	R01587 ye75ell_r1
C 38	6	37.5	107	9	AL651945 AL651945
C 39	6	37.5	135	9	AV060675 AV060675
C 40	6	37.5	143	10	AW582814 2f323gc
C 41	6	37.5	144	12	BG189371 RST8415 A
C 42	6	37.5	151	9	AA885939 o138a07.s
C 43	6	37.5	151	12	BE720567 OV4-HT089
C 44	6	37.5	152	10	AV626699 AV626699
C 45	6	37.5	152	10	AW794472 RC6-UM001
C 46	6	37.5	160	10	AW989089 AV626699
C 47	6	37.5	162	17	AZ357634 60099N15
C 48	6	37.5	167	10	BB422820 BB422820
C 49	6	37.5	175	9	AA873237 oh78e06.s
C 50	6	37.5	180	9	AV075220 AV075220
C 51	6	37.5	181	9	AI152761 ud94c08.r
C 52	6	37.5	183	10	BB575708 BB575708
C 53	6	37.5	196	10	AV360305 AV360305
C 54	6	37.5	199	9	AI156316 ue09c09.r
C 55	6	37.5	199	9	AV087856 AV087856
C 56	6	37.5	200	10	AW582852 gh09c2b29
C 57	6	37.5	200	13	B1189310 d4allfs.r
C 58	6	37.5	200	17	AZ386368 LM0145KL16
C 59	6	37.5	200	17	AZ386368 LM0145KL16
C 60	6	37.5	200	17	AZ716460 RPC1-24-1

ALIGNMENTS

REFERENCE 1 (bases 1 to 240)  
AUTHORS Urushihara,H.  
TITLE Developmental cDNA in Dictyostellium discoideum (1999)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan  
Tel: 81-298-53-4664  
Fax: 81-298-53-6614  
Email: hideko@biol.tsukuba.ac.jp  
PROJECT = Dictyostellium discoideum cDNA project in Japan.  
FEATURES  
source Location/Qualifiers  
1..240  
/organism="Dictyostellium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SSH359"  
/clone\_lib="Dictyostellium discoideum SS (H.Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 151 a 16 c 14 g 59 t  
ORIGIN

Alignment Scores:  
Pred. No.: 28.1 Length: 240  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AU073495 (1-240)

QY 9 PheillelePheTrpIlePhe 15  
|||||TTTTTTTTTTTTTTTT  
Db 50 TTTATTATTTCTGGATTTT 30

RESULT 3  
LOCUS BB261269/c  
DEFINITION musculus CDNA full-length enriched, 7 days neonate cerebellum Mus musculus A73009B10 3' similar to AL049783 Novel human gene mapping to chromosome 13, mRNA sequence.  
ACCESSION BB261269.1 GI:8957721  
VERSION BB261269.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 284)  
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kal,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shingagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toyota,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshinaka,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp.

TITLE BB261269/c  
JOURNAL  
COMMENT



URL:<http://genome.gsc.riken.go.jp/>  
Carninci, P., Nishiyama, Y., Muramatsu, M. and Hayashizaki, Y., N. Okazaki, Y., Muramatsu, M. and Hayashizaki, Y., Carninci, P. and Hayashizaki, Y., High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp/>) for further details.

## FEATURES

```

Location/Qualifiers
1. .284
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730098B10"
/clone.lib="RIKEN full-length enriched, 7 days neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/notes="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTTTTTNN 3'], cDNA was
prepared by using trenalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAATTAATCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
phage I."

```

BASE COUNT	94 a	70 c	60 g	60 t
ORIGIN				

Alignment Scores:	
Pred. No.:	32.1
Score:	7.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	43.75%
Gaps:	10
Indels:	0
Mismatches:	0
Conservative:	0
Matches:	7
Length:	284

MS-09-854-133-587 (1-16) X BB261269 (1-284)

Qy 5 CysGlyIleAspPheIleIle 11  
 |||||  
 Db 188 TGTGGTATAGATTTCATCATC 168

## RESULT 4

LOCUS	BE954331
DEFINITION	UI-M-CE1-bak-h-05-0-UI.s
	UI-M-CE1-bak-h-05-0-UI.s

ACCESSION BE954331.1  
VERSION BE954331.1  
KEYWORDS EST.  
GI:10596750

ORGANISM	Mus musculus
Eukaryota; Metazoa; Cho	
Mammalia; Eutheria; Rod	
REFERENCE	1 (bases 1 to 336)

AUTHORS	Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	<p>Contact: Chin, H          National Institute of Mental Health          6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD          20892-9643, USA          Tel: 301 443 1706          Fax: 301 443 9890          Email: mEST@mail.nih.gov</p> <p>Oligo-<math>\gamma</math>T track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements. The tissue for this library was contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine          Seq primer: M13 Forward          POLYA=NO.</p>

## FEATURES

FEATURES	source
Location/Qualifiers	
1. 336	
/organism="Mus musculus"	
/strain="C57BL/6J"	
/db_xref="taxon:10090"	
/clone="UI-M-CE1-bak-h-05-0-UI"	
/clone_lib="NIH_BMAP_Ret3_N"	
/dev_stage="6 weeks"	
/lab_host="DH10B (Life Technologies)"	
/note=vector: pW73D-Pac (Pharmacia)	
polylinker: Site.1: Not I; Site.2: Eco	
NIH_BMAP_Ret3_N library is a normalized	
NIH_BMAP_Ret3. NIH_BMAP_Ret3 was made	
tissue. For a detailed description of	
which this clone was derived, please	
brainerst.eng.ualowa.edu. The tissue for	
contributed by Dr. Xin-Yuan Fu, Yale	
Medicine	

BASE COUNT	92 a	62 c	82 g	100 t
------------	------	------	------	-------

Alignment Scores:	36.7	Length:	336
Pred. No.:	7.00	Matches:	7
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	43.75%	Gaps:	0

MS-00-054-133-587 (1-16) x BE954331 (1-336)

Qv 8 Asp phe Ile Ile phe phe Trp Ile 14

Db 49 GATTTCATCATATTCTGGATA 29

## RESULT 5

LOCUS BG446254

arborescens cDNA clone GA

VERSION BG446254.1 .GI:1335596

SOURCE Gossypium arboreum.

Eukaryota; Viridiplantae  
 Sporophyta: Magnoli

ROSIDAE; EURIOSUS II,  
 1 (bases 1 to 438)  
 PREFERENCE

[illegible]

ACCESSION BG446254  
 VERSION BG446254.1  
 KEYWORDS EST.  
 SOURCE Gossypium arboreum.

DIFFERENCE  
1 (bases 1 to 438)  
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Gossypium and relatives

## AUTHORS

Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
An integrated analysis of the genetics, development, and evolution of the cotton fiber  
Unpublished (2000)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu

## JOURNAL COMMENT

Seq primer: TAATACGACTCCTATAGG  
High quality sequence start: 6  
High quality sequence stop: 438.  
Location/Qualifiers  
1..438  
/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone\_lib="GA\_EB0033P23f"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="Fibers isolated from bolls harvested 7-10 dpa"  
/lab\_host="E. coli"  
/note="Vector: PBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

## FEATURES

## source

## Alignment Scores:

Pred. No.: 45.5 Length: 440  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BG446254 (1-438)

BASE COUNT 158 a 61 c 95 g 124 t  
ORIGIN

## Alignment Scores:

Pred. No.: 45.3 Length: 438  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BG446254 (1-438)

QY 10 IlellePheTrpIlePheTrp 16  
|||||  
Db 308 ATTATATTGGATATTGG 328

## RESULT 6

## BG446243

## LOCUS

DEFINITION GA\_EB0033N21f Gossypium arboreum 7-10 dpa fiber library Gossypium  
arborescens cDNA clone GA\_EB0033N21f, mRNA sequence.  
ACCESSION BG446243.1 GI:13355955  
VERSION BG446243.1  
KEYWORDS EST.  
SOURCE Gossypium arboreum.

## ORGANISM

## REFERENCE

1 (bases 1 to 440)  
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
An integrated analysis of the genetics, development, and evolution of the cotton fiber  
Unpublished (2000)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

Location/Qualifiers  
1..450  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2374E12"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"

## BASE COUNT

## ORIGIN

120 a 62 c 71 g 196 t 1 others

## FEATURES

## source

Location/Qualifiers  
1..440  
/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_EB0033N21f"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="Fibers isolated from bolls harvested 7-10 dpa"  
/lab\_host="E. coli"  
/note="Vector: PBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

## BASE COUNT

## ORIGIN

## Alignment Scores:

Pred. No.: 45.5 Length: 440  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BG446243 (1-440)

QY 10 IlellePheTrpIlePheTrp 16  
|||||  
Db 310 ATTATATTGGATATTGG 330

## RESULT 7

## AQ115066

## LOCUS

DEFINITION CIT-HSP-2374E12.TF CIT-HSP Homo sapiens genomic clone 2374E12, DNA  
sequence.  
ACCESSION AQ115066  
VERSION AQ115066.1 GI:3491187  
KEYWORDS GSS.  
SOURCE human.

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Use of a random human BAC End Sequence Database for Sequence-Ready  
Map Building  
Unpublished (1998)  
Other\_GSSs: CIT-HSP-2374E12.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13-21  
Class: BAC ends.

## FEATURES

## source

Location/Qualifiers  
1..450  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2374E12"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"

120 a 62 c 71 g 196 t 1 others

```

Alignment Scores:          46.3          Length:          450
Pred. No.:                7.00          Matches:          7
Score:                    100.00%        Conservative:      0
Percent Similarity:       100.00%        Mismatches:      0
Best Local Similarity:    100.00%        Indels:          0
Query Match:             43.75%         Gaps:           0
DB:                       17

US-09-854-133-587 (1-16) x AQ115066 (1-450)

Qy  9 PheillePheTrpIlePhe 15
    |||||||
Db  424 TTTATCATATTTTGGATATT 444

RESULT 8
LOCUS AU270164/c          452 bp      mRNA      linear      EST 10-MAY-2002
DEFINITION AU270164 VS Dictyostelium discoideum cDNA clone VSJ588 3', mRNA
ACCESSION AU270164
VERSION AU270164.1 GI:20528962
KEYWORDS EST
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
AUTHORS Urushihara,H., Morio,T., Saito,T., Koriki,E., Ochiai,H., Maeda,M.,
Takeuchi,I., Kohara,Y. and Tanaka,Y.
TITLE Population analysis of cDNAs from unicellular and multicellular
stages of Dictyostelium discoideum
JOURNAL Unpublished (2002)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp.

FEATURES
source
    Location/Qualifiers
        1..452
        /organism="Dictyostelium discoideum"
        /strain="AX4"
        /db_xref="taxon:44689"
        /clone="VSJ588"
        /clone_lib="VS"
        /sex="mat A"
        /dev_stage="vegetative"

BASE COUNT 184 a 60 c 78 g 130 t
ORIGIN
Alignment Scores:          46.5          Length:          452
Pred. No.:                7.00          Matches:          7
Score:                    100.00%        Conservative:      0
Percent Similarity:       100.00%        Mismatches:      0
Best Local Similarity:    100.00%        Indels:          0
Query Match:             43.75%         Gaps:           0
DB:                       9

US-09-854-133-587 (1-16) x AU270164 (1-452)

Qy  9 PheillePheTrpIlePhe 15
    |||||||
Db  164 TTTATAATTTTGGATCTTT 144

RESULT 9
LOCUS AU269237/c          457 bp      mRNA      linear      EST 10-MAY-2002
DEFINITION AU269237 VS Dictyostelium discoideum cDNA clone VSJ791 3', mRNA
ACCESSION AU269237
VERSION AU269237.1 GI:20528035
KEYWORDS EST

```

```

SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
AUTHORS 1 (bases 1 to 457)
Urushihara,H., Morio,T., Saito,T., Koriki,E., Ochiai,H., Maeda,M.,
Takeuchi,I., Kohara,Y. and Tanaka,Y.
TITLE Population analysis of cDNAs from unicellular and multicellular
stages of Dictyostelium discoideum
JOURNAL Unpublished (2002)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp.

FEATURES
source
    Location/Qualifiers
        1..457
        /organism="Dictyostelium discoideum"
        /strain="AX4"
        /db_xref="taxon:44689"
        /clone="VSJ791"
        /clone_lib="VS"
        /sex="mat A"
        /dev_stage="vegetative"

BASE COUNT 187 a 69 c 71 g 129 t 1 others
ORIGIN
Alignment Scores:          46.9          Length:          457
Pred. No.:                7.00          Matches:          7
Score:                    100.00%        Conservative:      0
Percent Similarity:       100.00%        Mismatches:      0
Best Local Similarity:    100.00%        Indels:          0
Query Match:             43.75%         Gaps:           0
DB:                       9

US-09-854-133-587 (1-16) x AU269237 (1-457)

Qy  9 PheillePheTrpIlePhe 15
    |||||||
Db  309 TTTATAATTTTGGATCTTT 289

RESULT 10
LOCUS BG543829
DEFINITION BG543829 Chinese cabbage etiolated seedling library Brassica rapa
subsp. pekinensis cDNA clone E1546, mRNA sequence.
ACCESSION BG543829
VERSION BG543829.1 GI:20374809
KEYWORDS EST.
SOURCE Brassica rapa subsp. pekinensis.
ORGANISM Brassica rapa subsp. pekinensis.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Eukaryota; Magnoliophyta; eudicotyledons; core eudicots;
Spermatophyta; Magnoliophyta; Brassicales; Brassicaceae; Brassica.
TITLE Rosidae; eurosids II;
JOURNAL 1 (bases 1 to 460)
COMMENT Ryu,S.H., Kang,J.S., Kang,C.-h., Kim,C.Y., Choi,Y.J., Lee,S.-H.,
Bahk,J.D., Lee,S.Y., Cho,M.J. and Lim,C.O.
Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
Unpublished (2001)
Contact: Lim, C.O.
Plant Molecular Biology & Biotechnology Research Centre
Gyeongsang National University
#900 Gazwa-dong, Jinju 660-701, Korea
Tel: 82 55 751 6255
Fax: 82 55 759 9363
Email: colim@nongae.gsnu.ac.kr
Seq primer: T7.
Location/Qualifiers
    1..460
    /organism="Brassica rapa subsp. pekinensis"
    /cultivar="Jangwon"
    /db_xref="taxon:51351"

```

```

/clone="E1546"
/tissue_type="Chinese cabbage etiolated seedling library"
/lab_host="XL-1 Blue"
/notes=vector: pSPORT 1; Site_1: Sal I; Site_2: Not I
BASE COUNT 137 a 93 c 87 g 143 t
ORIGIN

Alignment Scores:
Pred. No.: 47.2 Length: 460
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BG543829 (1-460)
Qy 9 PhelellePheTrpIlePhe 15
|||||
Db 127 TTTATAATCTTTGCGATTTT 147

RESULT 11
A2246141/c
LOCUS A2246141 465 bp DNA linear GSS 15-JUN-2000
DEFINITION RPCI-23-38C13.TJB RPCI-23 Mus musculus genomic clone RPCI-23-38C13,
DNA sequence.
ACCESSION A2246141
VERSION A2246141.1 GI:8559432
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 465)
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret,
and Fraser, C.M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-38C13.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 38 row: C column: 13
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1..465
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-38C13"
/db_xref="taxon:10090"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

```

```

BASE COUNT 236 a 60 c 65 g 104 t
ORIGIN

Alignment Scores:
Pred. No.: 47.6 Length: 465
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x AZ246141 (1-465)
Qy 9 PhelellePheTrpIlePhe 15
|||||
Db 294 TTTATAATCTTTGCGATTTT 274

RESULT 12
BF356741/c
LOCUS BF356741 506 bp mRNA linear EST 22-NOV-2000
DEFINITION PM3-HT0909-130900-002-h06 HT0909 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF356741
VERSION BF356741.1 GI:11315815
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 506)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&t2=PM3-HT0909-
130900-002-h06&t3=2000-09-13&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 12
High quality sequence stop: 505.

FEATURES
source
1..506
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0909"
/dev_stage="Adult"
/notes="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 165 a 98 c 106 g 137 t
ORIGIN

Alignment Scores:
Pred. No.: 50.9 Length: 506
Score: 7.00 Matches: 7

```

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BF356741 (1-506)

QY 6 GlytIleAspPheIleIlePhe 12  
 Db 300 GGTATTGATTTTATCATTTTC 280

RESULT 13

LOCUS BG097509

DEFINITION EST461944 potato leaves and petioles Solanum tuberosum cDNA clone  
 CSTB47G13 5' sequence, mRNA sequence.

ACCESSION BG097509

VERSION BG097509.1

KEYWORDS EST.

SOURCE potato.

ORGANISM

Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridiales; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 511)

AUTHORS van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,

Uterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning

,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.

Generation of ESTs from potato leaves and petioles

Unpublished (2000)

CONTACT: Cathy Ronning

The Institute for Genomic Research

For clone info: please contact Research Genetics, Libraries

Division tel 1-800-711-6195, email cdna@resgen.com.

LOCATION/Qualifiers

1. 511

/organism="Solanum tuberosum"

/cultivar="Kennebec"

/db\_xref="taxon:4113"

/clones="CSTB47G13"

/clone\_lib="potato leaves and petioles"

/tissue\_type="leaflets and petioles"

/dev\_stage="8 weeks old plants"

/lab\_host="SOLR"

/note="vector: pBluescript SK(-); Site.1: EcoR1; Site.2:

Xho1; Tissue was supplied by Dr. Fry (Cornell University).

Leaflets and petioles were isolated from 8 week old

greenhouse grown plants. The plants were watered and

fertilized freely. The tissue was immediately frozen in

liquid nitrogen."

BASE COUNT 153 a 64 c 112 g 182 t

ORIGIN

Alignment Scores:

Pred. No.: 51.3 Length: 511

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0

DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BG097509 (1-511)

QY 3 AlaAsnCysGlyIleAspPhe 9

Db 420 GCRAATTGTGATTTATTTT 440

RESULT 14

LOCUS BJ528250

DEFINITION BJ528250 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB032A10 5',

mRNA sequence.

ACCESSION BJ528250

VERSION BJ528250.1

KEYWORDS EST.

SOURCE Japanese medaka.

ORGANISM

Oryzias latipes

REFERENCE 1 (bases 1 to 529)

AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.

TITLE Medaka EST Project in Takeda's lab

JOURNAL Unpublished (2001)

COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES

source

1. 529

/organism="Oryzias latipes"

/strain="Hd-IR"

/db\_xref="taxon:8090"

/clone="MF01SSB032A10"

/clone\_lib="MF01SSB cDNA"

/sex="mixture of female and male"

/tissue\_type="whole embryo"

/dev\_stage="segmentation stage 20 - 25"

BASE COUNT 147 a 83 c 83 g 214 t

ORIGIN

Alignment Scores:

Pred. No.: 52.7 Length: 529

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0

DB: 13 Gaps: 0

US-09-854-133-587 (1-16) x BJ528250 (1-529)

QY 9 PheIleIlePheTrpIlePhe 15

Db 488 TTTATCATATTTTGGATTTT 468

RESULT 15

LOCUS BE971441/c

DEFINITION 601651432F1 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:3935070 5',

mRNA sequence.

ACCESSION BE971441

VERSION BE971441.1

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 572)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs@email.nih.gov

Tissue procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM776 row: 0 column: 07

High quality sequence stop: 571.

FEATURES  
Source

Location/Qualifiers  
1..572  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3935070"  
/clone\_lib="NIH\_MGC\_81"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCGGCACATG-dT(30)BN-3' (where B = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 235 a 109 c 104 g 124 t  
ORIGIN

## Alignment Scores:

Pred. No.: 56.1 Length: 572  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BE971441 (1-572)

QY 9 PheillePheTriPhe 15

Db 513 TTCATTATTTTGGATT 493

RESULT 16

BG571089/c

LOCUS

DEFINITION 602591696F1 NIH\_MGC\_79 Homo sapiens cDNA clone IMAGE:4714247 5', mRNA sequence.

ACCESSION BG571089

VERSION BG571089.1 GI:13578742

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 598)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapps@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI558 row: p column: 24

High quality sequence stop: 598.

FEATURES

Source

Location/Qualifiers

1..598

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4714247"

/clone\_lib="NIH\_MGC\_79"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: placenta; Vector: pDNR-LIB (Clontech);

Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattggcc)

; 5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGGCGGCGGCACATG-dT(30)BN-3' (where B = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

BASE COUNT 188 a 107 c 144 g 159 t  
ORIGIN

## Alignment Scores:

Pred. No.: 58.1 Length: 598  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BG571089 (1-598)

QY 6 GlyileAspPheillePhe 12

Db 28 GGGATTGACTTCATCATTT 8

RESULT 17

BH539329/c

LOCUS

DEFINITION BH539329 612 bp DNA linear GSS 14-DEC-2001

ACCESSION BH539329

VERSION BH539329.1 GI:17782077

KEYWORDS GSS.

SOURCE

ORGANISM

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 612)

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other\_GSSs: BOGMZ37TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES

source

Location/Qualifiers

1..612

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOGMZ37"

/clone\_lib="BOGM"

/note="Vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHOS1 using BstXI linkers"

BASE COUNT 195 a 116 c 150 g 151 t

ORIGIN

## Alignment Scores:

Pred. No.: 59.2 Length: 612  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH539329 (1-612)

Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x AZ316177 (1-617)

QY 9 PherleillePherTrpIlePhe 15  
Db 595 TTTATCAATCTTCTGGATTTTC 615

RESULT 19  
LOCUS AZ937525/c  
DEFINITION 2M0195C17R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
ACCESSION AZ937525  
VERSION AZ937525.1 GI:13796433  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 620)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
Plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0195 row: C column: 17  
Seq primer: CACACAGGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 620.  
Location/Qualifiers  
1. .620  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0195C17"  
/clone\_lib="Mouse 10kb plasmid UUGC2M library"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (female) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gi|4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 210 a 127 c 82 g 201 t

ORIGIN

QY 9 PherleillePherTrpIlePhe 15  
Db 595 TTTATCAATCTTCTGGATTTTC 615

RESULT 19  
LOCUS AZ937525/c  
DEFINITION 2M0195C17R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
ACCESSION AZ937525  
VERSION AZ937525.1 GI:13796433  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 617)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
Plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0033 row: A column: 16  
Seq primer: CACACAGGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 617.  
Location/Qualifiers  
1. .617  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0033A16"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gi|4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 195 a 119 c 129 g 173 t 1 others

ORIGIN

Alignment Scores:  
Pred. No.: 59, 6 Length: 617  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

## Alignment Scores:

Pred. No.: 59.8 Length: 620  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x AZ937525 (1-620)

QY 2 GlnAaenCysGlyleasp 8

Db 395 CAAGCAATGTGGCATTGAT 375

RESULT 20

BU008061

LOCUS

DEFINITION

QGH6F17.yg.abl OG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone

ACCESSION

BU008061

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..642

/organism="Lactuca sativa"

/cultivar="L.serriola"

/db\_xref="taxon:4236"

/clone="QGH6F17"

/lab\_host="E.coli"

/note="Vector: pBRCDNASFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformationally made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/ TAG\_LIB-OG\_EFGHJ lettuce serriola TAG\_TISSUE=germinating seeds TAG\_SEQ=TCGTGCGGG"

BASE COUNT 201 a 113 c 162 g 165 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 61.5 Length: 642

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 14 Gaps: 0

US-09-854-133-587 (1-16) x BU008061 (1-642)

QY 6 GlyleaspPhelellelepe 12

Db 444 GGGATGATTTATCATCTTC 464

RESULT 21

BU008061

LOCUS

DEFINITION

QGH6F17.yg.abl OG\_ABCDI lettuce salinas Lactuca sativa cDNA clone

ACCESSION

BU008061

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..675

/organism="Lactuca sativa"

/cultivar="Salinas"

/db\_xref="taxon:4236"

/clone="QGH6F17"

/lab\_host="E.coli"

/note="Vector: pBRCDNASFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformationally made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/ TAG\_LIB-OG\_ABCDI lettuce salinas TAG\_TISSUE=germinating seeds TAG\_SEQ=TCGTGCGGG"

BASE COUNT 215 a 116 c 172 g 172 t

ORIGIN

Alignment Scores:

Pred. No.: 64 Length: 675

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0

DB: 14 Gaps: 0



US-09-854-133-587 (1-16) x BQ868209 (1-675)

QY 6 GlyIleAspPheIleIlePhe 12  
|||||  
Db 438 GGGATTGATTTTATCACTTC 458

## RESULT 22

AUI36235 710 bp mRNA linear EST 02-AUG-2002  
LOCUS AUI36235 PLACE1 Homo sapiens cDNA clone PLACE1003825 5', mRNA  
DEFINITION

sequence.

ACCESSION AUI36235

VERSION AUI36235.1 GI:10996774

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 710)

Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,

Yanamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and

Isogai, T.

HRI human cDNA project

Unpublished (2000)

Contact: Takao Isogai

Genomics Laboratory

Helix Research Institute

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3975

Fax: 81-438-52-3986

Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix

Research Institute; cDNA library construction: Department of

Virology, Institute of Medical Science, University of Tokyo, and

Helix Research Institute.

FEATURES

source

1. 710

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="PLACE1003825"

/clone\_lib="PLACE1"

/tissue\_type="placenta"

/note="Vector: pME18SFL3"

BASE COUNT 206 a 150 c 104 g 244 t 6 others

ORIGIN

Alignment Scores:

Pred. No.: 66.7 Length: 710

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0

DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x AUI36235 (1-710)

QY 1 PheGlnAlaAsnCysGlyIle 7

|||||

Db 506 TTCAAGCCAACTGTGGGATA 526

RESULT 23

BH695614 717 bp DNA linear GSS 20-FEB-2002

LOCUS BH695614 BOHZ214TF BO\_2\_3\_KB Brassica oleracea genomic clone BOHZ214, DNA

DEFINITION

sequence.

ACCESSION BH695614

VERSION BH695614.1 GI:18768230

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 717)

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other\_GSSs: BOHZ214TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1. 717

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOHZ214"

/clone\_lib="BO\_2\_3\_KB"

/note="Vector: pPHOS1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into pPHOS1 using BstXI linkers"

BASE COUNT 257 a 138 c 113 g 209 t

ORIGIN

Alignment Scores:

Pred. No.: 67.2 Length: 717

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0

DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH695614 (1-717)

QY 6 GlyIleAspPheIleIlePhe 12

|||||

Db 472 GGTATAGATTTTATATATT 492

RESULT 24

BH273281 747 bp DNA linear GSS 30-NOV-2001

LOCUS BH273281 CH230-23H8 TV CHORI-230 Segment 1 Rattus norvegicus genomic clone

DEFINITION

CH230-23H8, DNA sequence.

ACCESSION BH273281

VERSION BH273281.1 GI:17185683

KEYWORDS GSS.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 747)

Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn

A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de

Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Other\_GSSs: CH230-23H8.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/orering-information.htm). BAC end

page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html

Plate: 23 row: H column: 8  
Seq primer: T7  
Class: BAC ends.

## FEATURES

source  
1. .747  
/organism="Rattus norvegicus"  
/strain="BN/SSNHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-23H8"  
/clone\_lib="CHORI-230 Segment 1"  
/sex="Female"  
/cell\_type="Brain"  
/note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;  
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
Pieter de Jong"

BASE COUNT 192 a 178 c 156 g 221 t  
ORIGIN

## Alignment Scores:

Pred. No.: 69.4 Length: 747  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH273281 (1-747)

QY 9 PheillePheTrpIlePhe 15

Db 390 TTTATCATTTCTCGGATCTTT 410

## RESULT 25

BH422412  
LOCUS  
DEFINITION BH422412 753 bp DNA linear GSS 12-DEC-2001  
BOHHD57TR BOHH Brassica oleracea genomic clone BOHHD57, DNA  
sequence.

ACCESSION BH422412

VERSION BH422412.1 GI:17608140

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 753)

AUTHORS Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished (2001)

COMMENT Other\_GSSs: BOHHD57TF

Contact: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES

source

1. .753 Location/Qualifiers

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOHHD57"

/clone\_lib="BOHH"

/note="Vector: pTARBAC2.1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 185 a 190 c 136 g 242 t

ORIGIN

Alignment Scores:

Pred. No.: 69.9 Length: 753

Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH422412 (1-753)

QY 9 PheillePheTrpIlePhe 15

Db 354 TTTATCATTTCTCGGATTTT 374

## RESULT 26

BH566426/c

LOCUS

DEFINITION BH566426 794 bp DNA linear GSS 14-DEC-2001

BOHSG71TR BOHS Brassica oleracea genomic clone BOHSG71, DNA

sequence.

ACCESSION BH566426

VERSION BH566426.1 GI:17818266

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 794)

AUTHORS Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished (2001)

COMMENT Other\_GSSs: BOHSG71TF

Contact: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES

source

1. .794 Location/Qualifiers

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOHSG71"

/clone\_lib="BOHS"

/note="Vector: pTARBAC2.1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 297 a 102 c 109 g 286 t

ORIGIN

Alignment Scores:

Pred. No.: 72.9 Length: 794

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 43.75% Indels: 0

DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH566426 (1-794)

QY 10 lleillePheTrpIlePheTrp 16

Db 586 ATATTTTCTGGATATTTGG 566

## RESULT 27

AQ898772

LOCUS

DEFINITION AQ898772 799 bp DNA linear GSS 10-NOV-1999

HS\_3130\_A2\_C05\_T7C CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=3130 Col=10 Row=E, DNA sequence.

ACCESSION AQ898772

VERSION AQ898772.1 GI:6354962

KEYWORDS GSS.

SOURCE  
ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 799)  
Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,  
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and  
Hood L.

## TITLE

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

## JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

## MEDLINE

99380589

## COMMENT

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 3130 row: E column: 10  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 799.

## FEATURES

source

1..799  
Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate-3130 Col-10 Row-E"

/clone.lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in

E-Coli DH10B"

233.a 171 c 141 g 253 t 1 others

BASE COUNT  
ORIGIN

## Alignment Scores:

Pred. No.: 73.2 Length: 799  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x A0898772 (1-799)

QY 10 IleIlePheTrpIlePheTrp 16

|||||

Db 498 ATTATCTCTGGATATTGG 518

## RESULT 28

BH152622

## LOCUS

BH152622 864 bp DNA linear GSS 27-AUG-2001

## DEFINITION

ENTPG74TR Entamoeba histolytica Sheared DNA Entamoeba histolytica

## ACCESSION

BH152622

## VERSION

BH152622.1 GI:15316561

## KEYWORDS

GSS.

## SOURCE

Entamoeba histolytica.

## ORGANISM

Entamoeba histolytica

## REFERENCE

1 (bases 1 to 864)

## AUTHORS

Lofthus B., Wang Z., Van Aken S. and Fraser C.

## TITLE

Determination of clone end sequences from Entamoeba histolytica

## JOURNAL

HMI:IMSS sheared DNA library (2001)

## COMMENT

Unpublished (2001)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

## Email: bjoftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

## DNA library

Seq primer: M13-Reverse

## Class: shotgun

High quality sequence start: 20

High quality sequence stop: 829.

## FEATURES

source

Location/Qualifiers

1..864

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db\_xref="taxon:5759"

/clone.lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOS1; Site:1; Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999)."

BASE COUNT 328 a 102 c 115 g 319 t

## ORIGIN

## Alignment Scores:

Pred. No.: 78 Length: 864  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH152622 (1-864)

QY 1 PheGlnAlaAsnCysGlyTle 7

|||||

Db 84 TTTTCAGGCTAATTCGGTATT 104

## RESULT 29

AZ692543

## LOCUS

AZ692543 879 bp DNA linear GSS 14-DEC-2000

## DEFINITION

ENTW09TR Entamoeba histolytica Sheared DNA Entamoeba histolytica

## ACCESSION

AZ692543

## VERSION

AZ692543.1 GI:11829689

## KEYWORDS

GSS.

## SOURCE

Entamoeba histolytica.

## ORGANISM

Entamoeba histolytica

## REFERENCE

1 (bases 1 to 879)

## AUTHORS

Lofthus B., Van Aken S. and Fraser C.

## TITLE

Determination of clone end sequences from Entamoeba histolytica

## JOURNAL

HMI:IMSS sheared DNA library

## COMMENT

Unpublished (2000)

## CONTACT

Contact: Brendan J Loftus

## DEPARTMENT

Department of Eukaryotic Genomics

## INSTITUTE

The Institute for Genomic Research

## ADDRESS

9712 Medical Center Dr., Rockville, MD 20850, USA

## TEL

Tel: 301 838 0208

## FAX

Fax: 301 838 3543

## EMAIL

Email: bjoftus@tigr.org

## CLONES

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

## DNA library

DNA library

## Seq primer:

Seq primer: M13-Reverse

## Class:

Class: shotgun

## High quality sequence start:

High quality sequence start: 15

## High quality sequence stop:

High quality sequence stop: 754.

## Location/Qualifiers

## FEATURES

```

source
1. .879
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (-2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      314 a 110 c 130 g 325 t
ORIGIN
Alignment Scores:
Pred. No.:      79          Length:      879
Score:          7.00        Matches:      7
Percent Similarity: 100.00%  Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    43.75%      Indels:      0
DB:             17          Gaps:       0

US-09-854-133-587 (1-16) x AZ692543 (1-879)
QY 1 PheGlnAlaAsnCysGlyIle 7
|||||
Db 509 TTTCAGGCCTAATGCCGTATT 529

RESULT 30
AZ692069
LOCUS
DEFINITION
ENTME42TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic DNA sequence.
ACCESSION
AZ692069
VERSION
AZ692069.1 GI:11829110
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica.
ORGANISM
Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
1 (bases 1 to 882)
Loftus,B., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 19
High quality sequence stop: 824.
Location/Qualifiers
1. .882
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica

```

```

using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (-2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      293 a 134 c 124 g 331 t
ORIGIN
Alignment Scores:
Pred. No.:      79.2        Length:      882
Score:          7.00        Matches:      7
Percent Similarity: 100.00%  Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    43.75%      Indels:      0
DB:             17          Gaps:       0

US-09-854-133-587 (1-16) x AZ692069 (1-882)
QY 9 PheIlelePheTrpIlePhe 15
|||||
Db 804 TTCATTATATTTGGATT 824

RESULT 31
BF675705/C
LOCUS
DEFINITION
BF675705
BF675705.1 GI:11949600
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 883)
NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1065 row: h column: 04
High quality sequence stop: 608.
Location/Qualifiers
1. .883
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4247859"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggcgctcgcc); Site_2: SfiI (ggccattatggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCGGCGATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT      336 a 205 c 164 g 178 t
ORIGIN

```

## Alignment Scores:

Pred. No.: 79.3 Length: 883  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BF675705 (1-883)

QY 9 PheillelePheTrpIlePhe 15  
 |||||  
 Db 573 TTCATTATATTTGGATTTT 553

## RESULT 32

BH130713/c  
 LOCUS 908 bp DNA linear GSS 14-NOV-2000  
 DEFINITION Entg31TR Entamoeba histolytica sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION AZ545454  
 VERSION AZ545454.1 GI:11166245  
 KEYWORDS GSS.

SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica.

REFERENCE 1 (bases 1 to 908)  
 Eukaryota; Entamoebidae; Entamoeba.

AUTHORS Loftus, B., Van Aken, S. and Fraser, C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library  
 COMMENT Unpublished (2000)

CONTACT: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
 Seq primer: M13-Reverse  
 Class: shotgun  
 High quality sequence start: 99  
 High quality sequence stop: 526.

Location/Qualifiers

FEATURES  
 source  
 1..908  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: PHOS1; Site\_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 370 a 97 c 149 g 292 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 81.1 Length: 908  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH130713 (1-908)

QY 9 PheillelePheTrpIlePhe 15  
 |||||  
 Db 772 TTTATTATTTCTGGATATTC 752

## RESULT 34

BH130713/c  
 LOCUS 908 bp DNA linear GSS 07-AUG-2001  
 DEFINITION ENTX17TF Entamoeba histolytica sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION BH130713  
 VERSION BH130713.1 GI:15089062  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica.  
 REFERENCE 1 (bases 1 to 908)  
 Eukaryota; Entamoebidae; Entamoeba.  
 AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library (2001)  
 COMMENT Unpublished (2001)  
 CONTACT: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence stop: 663.  
 Location/Qualifiers

FEATURES  
 source  
 1..908  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: PHOS1; Site\_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

US-09-854-133-587 (1-16) x AZ545454 (1-908)

QY 9 PheillelePheTrpIlePhe 15  
 |||||  
 Db 658 TTCATAATATTTGGATCTTT 678

## RESULT 33

BH130713/c

LOCUS 908 bp DNA linear GSS 07-AUG-2001  
 DEFINITION ENTX17TF Entamoeba histolytica sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION BH130713  
 VERSION BH130713.1 GI:15089062  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica.

REFERENCE 1 (bases 1 to 908)  
 Eukaryota; Entamoebidae; Entamoeba.

AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library (2001)  
 COMMENT Unpublished (2001)  
 CONTACT: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence stop: 663.  
 Location/Qualifiers

FEATURES  
 source  
 1..908  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: PHOS1; Site\_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 352 a 116 c 152 g 288 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 81.1 Length: 908  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 43.75% Indels: 0  
 DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH130713 (1-908)

QY 9 PheillelePheTrpIlePhe 15  
 |||||  
 Db 772 TTTATTATTTCTGGATATTC 752

## RESULT 34

BH130713/c

LOCUS 908 bp DNA linear GSS 07-AUG-2001  
 DEFINITION ENTX17TF Entamoeba histolytica sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION BH130713  
 VERSION BH130713.1 GI:15089062  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica.

REFERENCE 1 (bases 1 to 908)  
 Eukaryota; Entamoebidae; Entamoeba.

AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library (2001)  
 COMMENT Unpublished (2001)  
 CONTACT: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence stop: 663.  
 Location/Qualifiers

FEATURES  
 source  
 1..908  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: PHOS1; Site\_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 352 a 116 c 152 g 288 t

## ORIGIN

```

LOCUS
DEFINITION 601846206F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4077432 5', linear EST 06-NOV-2000
mRNA sequence.
ACCESSION BI215221
VERSION BI215221.1 GI:11108807
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 971)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM932 row: k column: 01
High quality sequence start: 19
High quality sequence stop: 62.
FEATURES
Location/Qualifiers
1..971
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4077432"
/tissue_type="NIH_MGC_55"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgcttcggcc); Site_2: SfiI (ggccattatggcc)
; Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGCGCATTTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCGGCGCATG-dT(30)BN-3'
(where B = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 329 a 201 c 207 g 234 t
ORIGIN
Alignment Scores:
Pred. No.: 85.6 Length: 971
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 12 Gaps: 0
US-09-854-133-587 (1-16) x BF215221 (1-971)
QY 9 PhelellePheTrpillePhe 15
|||||
Db 657 TTCATTATATTGGATTATT 637
RESULT 35
BE250032
LOCUS 600943035F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:2959497 5', linear EST 13-JUL-2000
mRNA sequence.
ACCESSION BE250032
VERSION BE250032.1 GI:9120137
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 996)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM51 row: f column: 10
High quality sequence stop: 669.
FEATURES
Location/Qualifiers
1..996
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2959497"
/tissue_type="NIH_MGC_15"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTa7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 196 a 319 c 284 g 197 t
ORIGIN
Alignment Scores:
Pred. No.: 87.3 Length: 996
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 10 Gaps: 0
US-09-854-133-587 (1-16) x BE250032 (1-996)
QY 1 PheGlnAlaAsnCysGlyIle 7
|||||
Db 665 TTCAGGCGGAATTGTGGATC 685
RESULT 36
BI260864
LOCUS 602971086F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5110436 5', linear EST 17-JUL-2001
mRNA sequence.
ACCESSION BI260864
VERSION BI260864.1 GI:14819571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1077)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

```

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 996)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM51 row: f column: 10
High quality sequence stop: 669.
FEATURES
Location/Qualifiers
1..996
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2959497"
/tissue_type="NIH_MGC_15"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTa7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 196 a 319 c 284 g 197 t
ORIGIN
Alignment Scores:
Pred. No.: 87.3 Length: 996
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.75% Indels: 0
DB: 10 Gaps: 0
US-09-854-133-587 (1-16) x BE250032 (1-996)
QY 1 PheGlnAlaAsnCysGlyIle 7
|||||
Db 665 TTCAGGCGGAATTGTGGATC 685
RESULT 36
BI260864
LOCUS 602971086F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5110436 5', linear EST 17-JUL-2001
mRNA sequence.
ACCESSION BI260864
VERSION BI260864.1 GI:14819571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1077)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

Plate: LLAM11267 row: 1 column: 21

High quality sequence stop: 700.

## FEATURES

source

Location/Qualifiers

1. .1077

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5110436"

/clone\_lib="NIH\_MGC\_12"

/tissue\_type="cervical carcinoma cell line"

/lab\_host="DH10B"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.4 kb. Library prepared by Life

Technologies."

BASE COUNT 335 a 258 c 210 g 274 t

ORIGIN

Alignment Scores:

Pred. No.: 92.9 Length: 1077  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 43.75% Indels: 0  
DB: 13 Gaps: 0

US-09-854-133-587 (1-16) x B1260864 (1-1077)

Qy 9 PhelellePheTrpIlePhe 15

|||||

Db 138 TTCATAATCTTTGGATTTC 158

RESULT 37

R01587

LOCUS

DEFINITION R01587 91 bp mRNA linear EST 31-MAR-1995

ye75ell.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone

IMAGE:123596 5' similar to gb:J03779 NEPRILYSIN (HUMAN);, mRNA

sequence.

ACCESSION R01587.1 GI:751323

VERSION

KEYWORDS EST.

SOURCE human.

ORGANISM

REFERENCE

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston

, R., Williamson, A., Wohldmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilton RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1458

High quality sequence stops: 70 Source: IMAGE Consortium, LLNL This

clone is available royalty-free through LLNL; contact the IMAGE

Consortium (info@image.llnl.gov) for further information.

Insert Length: 1458 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 70.

Location/Qualifiers

1. .91

/organism="Homo sapiens"

/db\_xref="GDB:476141"

/db\_xref="taxon:9606"

/clone="IMAGE:123596"

/clone\_lib="Soares fetal liver spleen INFLS"

/sex="male"

/dev\_stage="20 week-post conception fetus"

FEATURES

source

/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGGAAGAATAATTAAGATCTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 29 a 16 c 11 g 33 t

ORIGIN

Alignment Scores:

Pred. No.: 192 Length: 91  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 14 Gaps: 0

US-09-854-133-587 (1-16) x R01587 (1-91)

Qy 1 PheGlnAlaAsnCysGly 6

|||||

Db 67 TTTCAGGCCAACTGTGGA 84

RESULT 38

AL651945

LOCUS

DEFINITION AL651945 XGC-gastrula silurana tropicalis cDNA clone Tgas046111 5',

mRNA sequence.

ACCESSION AL651945.1 GI:17662266

VERSION

KEYWORDS EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (10\_2001)

Unpublished (2001)

Contact: Huckle E

Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: Tgas046111.sp6

Sequencing primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

Location/Qualifiers

1. .107

/organism="Silurana tropicalis"

/db\_xref="taxon:8364"

/clone="Tgas046111"

/clone\_lib="XGC-gastrula"

/dev\_stage="gastrula (stages 10.5-13 mixed)"

/lab\_host="Escherichia coli XL1-blue"

/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA

was oligo dT primed from 5' of poly A+ RNA from stages

10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated

into pCS107 with EcoRI at the 5' end and NotI at the 3'

end."

BASE COUNT 42 a 18 c 13 g 34 t

ORIGIN

Alignment Scores:

Pred. No.: 219 Length: 107  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00%  
 Query Match: 37.50%  
 DB: 9  
 Mismatches: 0  
 Indels: 0  
 Gaps: 0

US-09-854-133-587 (1-16) x AL651945 (1-107)

QY 9 PheillelePheTrp1e 14  
 |||||  
 Db 85 TTTATTATTCTGGATC 102

RESULT 39

AV060675

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AV060675 135 bp mRNA linear EST 23-JUN-1999  
 AV060675 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA  
 clone 1810064A19, mRNA sequence.

AV060675

AV060675.1 GI:5160422

EST.

Mus musculus

house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 135)

Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,

Akai, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara

, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,

Kikuchi, N., Kojima, Y., Matsuyama, T., Mitsuhashi, H., Oda, H., Owa, C.,

Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara

, Y., Suzuki, H., Suzuki, H., Tatenos, M., Tomaru, Y., Tomimaga, N.,

Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,

Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

RIKEN Mouse ESTs

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp

Thermolabile and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (<http://genome.rtc.riken.go.jp>) for

further details.

Location/Qualifiers

1..135

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="1810064A19"

/clone\_lib="Mus musculus pancreas C57BL/6J adult"

/sex="male"

/tissue\_type="pancreas"

/dev\_stage="adult"

/note="Organ: mammary gland; Vector: pT7n3D-Pac (Pharmacia

) with a modified polylinker; Site: 1: Not I; Site 2: Eco

RI; 1st strand cDNA was primed with a Not I - oligo (dT)

primer [5']

TGTTACCATCTGAAGTGGAGCGCGGAATGTTTTTTTTTTTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT7n3 vector.

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M.Fatima

Bonaldo."

37 a 40 c 26 g 32 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 263

Length: 135

QY 6 GlytleaspPheillele 11

|||||

Db 24 GGCATTGACATTCATTATA 41

Score: 6.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 37.50%  
 DB: 9  
 Matches: 6  
 Conservative: 0  
 Mismatches: 0  
 Indels: 0  
 Gaps: 0

US-09-854-133-587 (1-16) x AV060675 (1-135)

QY 8 AspPheillelePheTrp 13

|||||

Db 61 GACTTCATCATTTTGG 78

RESULT 40

AW582814

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AW582814 143 bp mRNA linear EST 01-APR-2000  
 2fj323gc Neuronal Differentiation of the NT2/D1 cell line. Homo  
 sapiens cDNA 3' similar to EST, mRNA sequence.

AW582814

AW582814.1 GI:7382060

EST.

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 143)

Bevort, M.

Analysis of gene expression during neuronal differentiation of

NT2/D1 cells

Unpublished (2000)

Contact: Bevort M

Department of Growth and Reproduction GR-5064

Copenhagen University Hospital

Blegdamsvej 9, 2100 Copenhagen, Denmark

Tel: +45 35455081

Fax: +45 35456054

Email: maja@biobase.dk

The EST is up regulated, during neuronal differentiation of the

NT2/D1 cell line (replated fully differentiated neurones not

analysed).

PCR Primers

FORWARD: GAGCAGGAATTC

BACKWARD: AAGCTTTT

Seq primer: T7, CY5-TAATACGACTCACTATAGGCCC

High quality sequence stop: 143.

Location/Qualifiers

1..143

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Neuronal Differentiation of the NT2/D1 cell

line."

/cell\_line="NT2/D1"

/note="The EST is derived from direct sequencing of a

Differential Display fragment. Laboratory manuals are

available from <http://www.biobase.dk/~ddbse>"

BASE COUNT 42 a 30 c 25 g 43 t 3 others

ORIGIN

Alignment Scores:

Pred. No.: 276

Length: 143

Score: 6.00

Matches: 6

Percent Similarity: 100.00%

Conservative: 0

Best Local Similarity: 100.00%

Mismatches: 0

Query Match: 37.50%

Indels: 0

Gaps: 0

US-09-854-133-587 (1-16) x AW582814 (1-143)

QY 6 GlytleaspPheillele 11

|||||

Db 24 GGCATTGACATTCATTATA 41

RESULT 41

BG189371/c



```

LOCUS      BG189371              144 bp      mRNA      linear      EST 21-APR-2001
DEFINITION RST8415 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION  BG189371
VERSION    BG189371.1  GI:13711058
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 144)
AUTHORS   Harrington,J.J., Sherr,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
            Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
            ,E., Veloso,N., Klika,A., Hess,J., Cottrien,K., Lo,K., Offenbacher
            ,J., Danzig,J. and Ducar,M.
TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE   21227151
COMMENT   Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@atersys.com
            High quality sequence stop: 144.
FEATURES   Location/Qualifiers
            source          1..144
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="Athersys RAGE Library"
                        /cell_line="HT1080"
                        /note="See 'Creation of Genome-wide Protein Expression
                        Libraries using Random Activation of Gene Expression',
                        Nature Biotechnology, in press. Note that even though the
                        cell type indicated is HT1080, since a random activation
                        method was used, these sequence tags are not necessarily
                        expressed in HT1080 under normal circumstances."
BASE COUNT  54 a      21 c      19 g      50 t
ORIGIN
Alignment Scores:
Pred. No.:      277      Length:      144
Score:          6.00      Matches:      6
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    37.50%      Indels:      0
DB:             12      Gaps:      0

US-09-854-133-587 (1-16) x BG189371 (1-144)

QY      4  AsnCysGlyleAspPhe 9
Db      129  AATTGGGAATAGATT 112

RESULT 42
AA885939
LOCUS    o38a07.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1500564 3',
DEFINITION mRNA sequence.
ACCESSION  AA885939
VERSION    AA885939.1  GI:3001047
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 151)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.

```

```

Email: cgabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1241 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 59.
FEATURES   Location/Qualifiers
            source          1..151
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1500564"
                        /clone_lib="NCI_CGAP_Kid3"
                        /lab_host="DH10B"
                        /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                        a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                        strand cDNA was primed with a Not I - oligo(dt) primer,
                        double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not
                        I and Eco RI sites of the modified pT7T3 vector. mRNA
                        source: 2 pooled kidneys. Library went through one round
                        of normalization. Library constructed by Bento Soares and
                        M. Fatima Bonaldo."
BASE COUNT  48 a      30 c      28 g      45 t
ORIGIN
Alignment Scores:
Pred. No.:      288      Length:      151
Score:          6.00      Matches:      6
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    37.50%      Indels:      0
DB:             9      Gaps:      0

US-09-854-133-587 (1-16) x AA885939 (1-151)

QY      6  GlyileAspPheille 11
Db      37  GGCATTGACTTCATTATA 54

RESULT 43
BE720567/c
LOCUS    BE720567
DEFINITION QV4-HT0894-270700-318-e03 HT0894 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE720567
VERSION    BE720567.1  GI:10108832
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 151)
AUTHORS   Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
            Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed

```

# JOURNAL MEDLINE COMMENT

sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV4-HT0894-270  
700-318-e03et3-2000-07-27et4-1)  
Seq primer: puc 18 forward  
High quality sequence start: 6  
High quality sequence stop: 151.

## FEATURES source

1. .151  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0894"  
/dev\_stage="Adult"

/note="Organ: head neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 47 a 31 c 32 g 41 t

## ORIGIN

### Alignment Scores:

Pred. No.: 288 Length: 151  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 12 Gaps: 0

US-09-854-133-587 (1-16) x BE720567 (1-151)

QY 6 GlyTleAspPheIleile 11

Db 143 GGCATTGACTCAATCATT 126

## RESULT 44 AV626699/c

LOCUS AV626699 152 bp mRNA linear EST 15-DEC-2000  
DEFINITION Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
reinhardtii CDNA clone LCU013h08\_r 5', mRNA sequence.

## ACCESSION AV626699

VERSION AV626699.1 GI:10788979

## KEYWORDS EST.

## SOURCE ORGANISM

Chlamydomonas reinhardtii.

Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 152)

Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohnaya,K.,

Nakamura,Y., and Tabata,S.

Generation of expressed sequence tags from low-CO2 and high-CO2

adapted cells of Chlamydomonas reinhardtii

DNA Res. 7 (5), 305-307 (2000)

20539644

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

## FEATURES

Location/Qualifiers

## source

1. .152  
/organism="Chlamydomonas reinhardtii"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone\_lib="LCU013h08\_r"

/clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI; the cDNA library was constructed from cells cultured  
in a carbon stress acclimatized condition in which carbon  
dioxide concentration in the bubbling gas was changed from  
5% to 0.04%"

BASE COUNT 42 a 46 c 29 g 35 t

## ORIGIN

### Alignment Scores:

Pred. No.: 290 Length: 152  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.50% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AV626699 (1-152)

QY 3 AlaAsnCysGlyIleileAsp 8

Db 39 GCGCACTGTGTATCGAC 22

## RESULT 45

AW794472

## LOCUS

RC6-UM0014-210200-011-F04 UM0014 Homo sapiens cDNA, mRNA sequence.  
152 bp mRNA linear EST 16-MAY-2000

## ACCESSION AW794472

VERSION AW794472.1 GI:7846342

## KEYWORDS EST.

## SOURCE ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 152)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,

Brustein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC6-UM0014-210

200-011-F04et3-2000-02-21et4-1)

Seq primer: puc 18 forward

High quality sequence stop: 152.

Location/Qualifiers

1. .152

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="UM0014"

/dev\_stage="Adult"

/note="Organ: uterus; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196

us-09-854-133-587.oligo.rst

Tue May 13 12:13:00 2003

,716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 51 a 33 c 49 g 19 t

ORIGIN

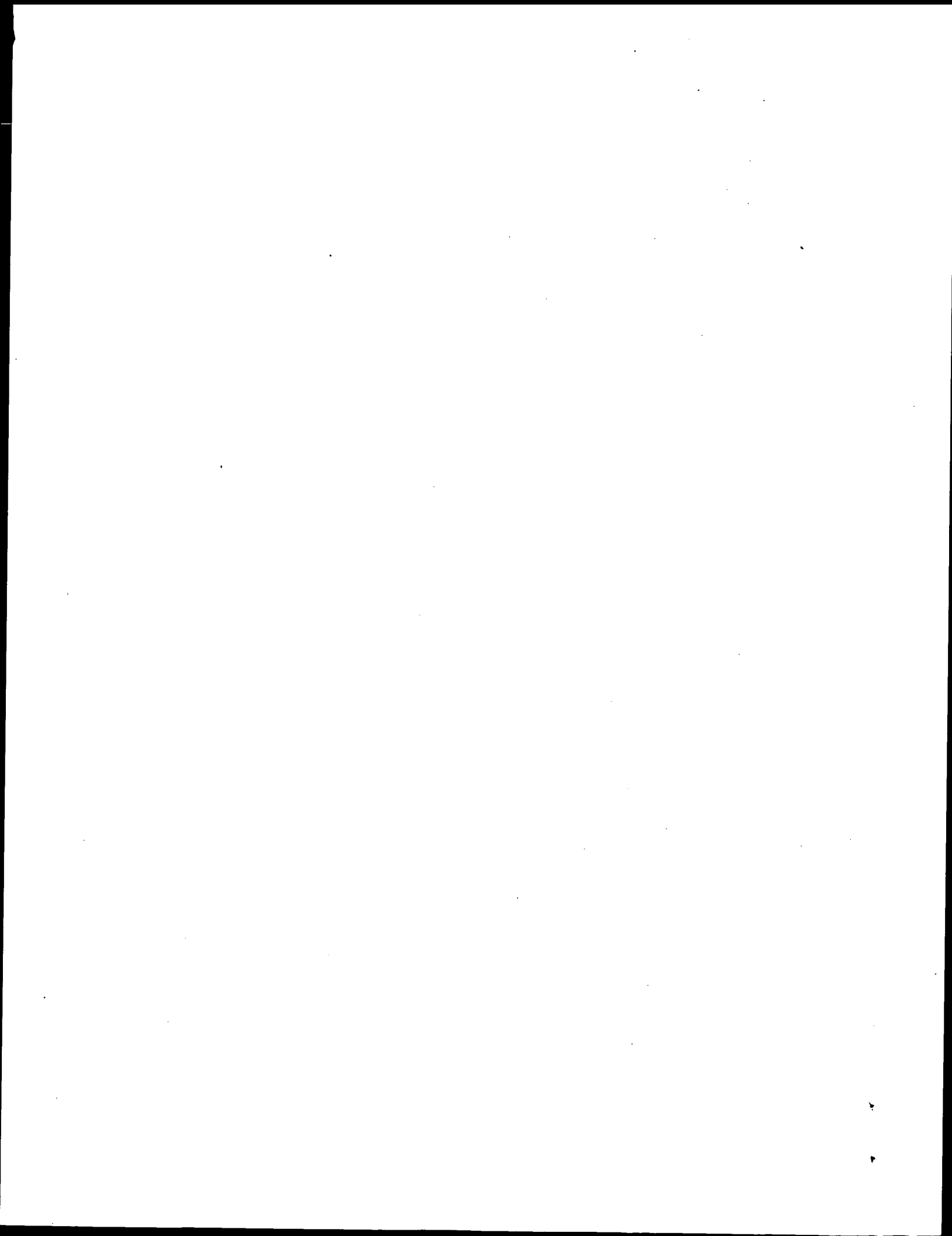
Alignment Scores: 290 Length: 152  
Pred. No.: 6.00 Matches: 6  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 37.50% Gaps: 0  
DB: 10

US-09-854-133-587 (1-16) x AW794472 (1-152)

QY 3 AlaAsnCysGlyIleAsp 8

Db 118 GCAAACTGTGGCATAGAC 135

Search completed: May 11, 2003, 17:46:22  
Job time : 249.973 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 19:32:42 ; Search time 9.34513 Seconds  
(without alignments)  
228.141 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FOAGCIDFIIFWIFW 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database :	A_Geneseq_101002.*
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	16	100.0	16	22	AAE13851	Human T cell epitope
2	6	37.5	399	17	AAW05523	HCMV Towne strain
3	5	31.2	44	22	ABB38939	Peptide #6445 enco
4	5	31.2	44	22	ABB23925	Protein #5924 enco
5	5	31.2	44	22	AAW59589	Human brain expres
6	5	31.2	44	22	AAW72162	Human bone marrow
7	5	31.2	44	22	AAW19471	Peptide #5905 enco
8	5	31.2	44	22	AAW32422	Peptide #6459 enco
9	5	31.2	44	23	ABG41976	Human peptide enco
10	5	31.2	62	21	AAB30672	Amino acid sequenc

# ALIGNMENTS

RESULT 1	AAE13851	standard; peptide; 16 AA.
ID	AAE13851	
XX	AAE13851;	
AC	AAE13851;	
XX	26-FEB-2002	(first entry)
DT	Human T cell epitope related to lung tumour-specific protein.	
XX	Human; lung tumour protein; immunostimulant; cytostatic; gene therapy; antisense-therapy; vaccine; immune response; lung cancer;	
KW	T cell epitope.	

11	5	31.2	62	21	AAB28390	Human CD28 transme
12	5	31.2	62	22	AAB84314	Amino acid sequenc
13	5	31.2	62	22	AAB98753	Human CD28 extrace
14	5	31.2	62	22	AAB98800	Human CD28 extrace
15	5	31.2	62	22	AAB28678	Protein encoded by
16	5	31.2	62	23	AAU98026	Human IgG1 hinge/C
17	5	31.2	62	22	AAW92938	Human digestive sy
18	5	31.2	67	22	AAU20034	Human liver associ
19	5	31.2	67	23	ABP40895	Human liver antige
20	5	31.2	68	21	AAB58985	Breast and ovarian
21	5	31.2	69	17	AAW94651	Mouse fas-associat
22	5	31.2	72	21	AAW00588	Human secreted pro
23	5	31.2	74	23	ABP40821	Staphylococcus epi
24	5	31.2	76	22	ABG27612	Novel human diago
25	5	31.2	81	18	AAW28161	Amino acid sequenc
26	5	31.2	82	22	AAW94589	Human reproductive
27	5	31.2	87	20	AAW60427	Human normal bladd
28	5	31.2	92	22	ABG27579	Novel human diago
29	5	31.2	92	22	AAW90743	Human immune/haema
30	5	31.2	103	20	AAW24470	Human CD28 gene pr
31	5	31.2	104	22	AAW94419	Human reproductive
32	5	31.2	111	22	ABG30130	Novel human diago
33	5	31.2	112	21	AAW40079	Arabidopsis thalia
34	5	31.2	118	23	ABW09650	Amino acid sequenc
35	5	31.2	120	23	ABW48192	Listeria monocytog
36	5	31.2	134	23	ABP40387	Staphylococcus epi
37	5	31.2	146	7	AAW60068	Sequence of HTLV-I
38	5	31.2	148	22	AAW92462	Human protein sequ
39	5	31.2	149	19	AAW85902	S. pneumoniae derl
40	5	31.2	150	18	AAW20434	H. pylori cytoplas
41	5	31.2	151	22	ABG30352	Novel human diago
42	5	31.2	152	22	ABG22986	Novel human diago
43	5	31.2	157	23	ABP29925	Streptococcus poly
44	5	31.2	157	23	ABP30763	Streptococcus poly
45	5	31.2	160	22	ABG12807	Novel human diago
46	5	31.2	166	22	AAU71909	C. glutamicum meta
47	5	31.2	166	22	AAU71910	C. glutamicum prote
48	5	31.2	166	22	AAW90777	Corynebacterium gl
49	5	31.2	166	22	AAW97941	Corynebacterium gl
50	5	31.2	166	22	AAW79886	Corynebacterium gl
51	5	31.2	166	22	AAW80088	Corynebacterium gl
52	5	31.2	166	22	AAW80069	Corynebacterium gl
53	5	31.2	170	8	AAW70262	Beta-glucuronidase
54	5	31.2	170	9	AAW80057	Beta-glucuronidase
55	5	31.2	170	15	AAW56657	Protein product of
56	5	31.2	195	23	ABW54459	Lactococcus lactis
57	5	31.2	195	23	ABW54459	Neisseria meningit
58	5	31.2	197	21	AAW75378	Neisseria gonorrhoe
59	5	31.2	198	21	AAW75377	Neisseria meningit
60	5	31.2	198	21	AAW75377	G protein-coupled
61	5	31.2	198	22	AAW25706	Mouse MrqC9 (mas-r
62	5	31.2	198	23	AAW21334	Human polypeptide
63	5	31.2	202	23	ABW89451	Human protein SEQ
64	5	31.2	205	22	AAW78783	Mouse MrqC7 (mas-r
65	5	31.2	206	23	AAE21332	

```

XX OS Homo sapiens.
XX PN WO200172295-A2.
XX PD 04-OCT-2001.
XX PF 28-MAR-2001; 2001WO-US09991.
XX PR 29-MAR-2000; 2000US-0538037.
XX PR 05-JUN-2000; 2000US-0588937.
XX PR 18-AUG-2000; 2000US-0640878.
XX PR 22-SEP-2000; 2000US-234517P.
XX PR 01-NOV-2000; 2000US-0704512.
XX PR 14-DEC-2000; 2000US-0738973.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Lodes M7, Mohamath R, Secrist H, Benson DR, Indirias CY;
XX PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX DR WPI; 2001-639201/73.
XX PT New human lung-specific polynucleotides and polypeptides for the
XX PT diagnosis and treatment of disease e.g. lung cancer -
XX PS Claim 2; Page 378; 378pp; English.
XX CC The invention relates to isolated lung tumour-specific proteins and
XX CC their corresponding cDNA molecules. Lung tumour-specific proteins and
XX CC their antigen-presenting cells are useful for stimulating and/or
XX CC expanding T cells specific for a tumour protein, and for inhibiting
XX CC the development of cancer. The invention also relates to a composition
XX CC useful for stimulating an immune response, and for treating cancer. The
XX CC lung tumour specific oligonucleotide is useful in gene therapy and for
XX CC diagnosis, detection and treatment of lung cancer. The present sequence
XX CC is human T cell epitope related to lung tumour-specific protein.
XX SQ Sequence 16 AA;

Query Match 100.0%; Score 16; DB 22; Length 16;
Best Local Similarity 100.0%; Pred. No. 9.8e-12;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANGGIDFIIFWIFW 16
Db 1 FOANGGIDFIIFWIFW 16

RESULT 2
AAW05523
ID AAW05523 standard; Protein; 399 AA.
XX AC AAW05523;
XX DT 15-JAN-1997 (first entry)
XX DE HCMV Towne strain UL154 protein.
XX DE HCMV; vaccine; diagnosis; UL154.
XX KW CMV; HCMV; vaccine; diagnosis; UL154.
XX OS Human cytomegalovirus strain Towne (ATCC VR 977).
XX PN WO9630387-A1.
XX PD 03-OCT-1996.
XX PF 26-MAR-1996; 96WO-US04100.
XX PR 31-MAR-1995; 95US-04114926.
XX PA (AVIR-) AVIRON.

```

```

PI Cha T, Spaete R;
XX WPI; 1996-455265/45.
XX DR N-PSDB; AAT41419.
XX PT New isolated human cytomegalovirus nucleic acid - from Towne and
XX PT Toledo strains, used to develop prods. for the diagnosis, prevention
XX PS and treatment of human CMV infections
XX PS Claim 5; Page 47-48; 150pp; English.
XX CC Novel protein UL154 (AAW05523) is the product of an open reading
XX CC frame found in a novel nucleic acid (AAT41419) isolated from the
XX CC Towne strain of human cytomegalovirus (HCMV). UL154 and other
XX CC novel proteins (see also AAW05521-22 and AAW05524) of the Towne strain,
XX CC as well as novel and known proteins (see also AAW05500-20) from HCMV
XX CC Toledo, can be produced in transformed host cells and used in the
XX CC prodn. of subunit vaccines against HCMV. They may be surface
XX CC glycoproteins that are immunogenic or responsible for tissue
XX CC tropism, or may influence the immune response of an infected
XX CC individual.
XX SQ Sequence 399 AA;

Query Match 37.5%; Score 6; DB 17; Length 399;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIFI 14
Db 20 FIIIFI 25

RESULT 3
ABB38939
ID ABB38939 standard; Peptide; 44 AA.
XX AC ABB38939;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #6445 encoded by human foetal liver single exon probe.
XX DE Human; foetal liver; gene expression; single exon nucleic acid probe.
XX KW Homo sapiens.
XX OS WO200157277-A2.
XX PN 09-AUG-2001.
XX PD 30-JAN-2001; 2001WO-US00669.
XX PF 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PN Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-483447/52.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PT Claim 27; SEQ ID NO 31574; 639pp + sequence listing; English.
XX PS The invention relates to a single exon nucleic acid probe for
XX CC

```

CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC fetal liver. The present sequence is a peptide encoded by a single exon  
 CC nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 44 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 GIDFI 10  
 Db |  
 9 GIDFI 13  
 RESULT 4  
 ABB23925  
 ID ABB23925 standard; Protein; 44 AA.  
 XX AC ABB23925;  
 XX DT 23-JAN-2002 (first entry)  
 XX DE Protein #5924 encoded by probe for measuring heart cell gene expression.  
 XX KW Human; gene expression; heart; microarray; vascular system;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease.  
 XX OS Homo sapiens.  
 XX PN WO200157274-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00666.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488899/53.  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 XX hearts -  
 PS Claim 15; SEQ ID No 25695; 530pp; English.  
 XX The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,  
 CC hypertension, cardiac arrhythmias and congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 44 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 GIDFI 10  
 Db |  
 9 GIDFI 13  
 RESULT 5  
 AAM59589  
 ID AAM59589 standard; Protein; 44 AA.  
 XX AC AAM59589;  
 XX DT 05-NOV-2001 (first entry)  
 XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31694.  
 XX KW Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer.  
 XX OS Homo sapiens.  
 XX PN WO200157275-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00667.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-483446/52.  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 XX brains -  
 PS Example 4; SEQ ID NO: 31694; 650pp + Sequence Listing; English.  
 XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention.  
 XX SQ Sequence 44 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 GIDFI 10  
 Db |  
 9 GIDFI 13  
 RESULT 6

AAM72162  
 ID AAM72162 standard; Protein; 44 AA.  
 AC AAM72162;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32468.  
 DE  
 DE Human bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200157276-A2.  
 PN  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00568.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PA  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI  
 XX WPI; 2001-488900/53.  
 DR  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX  
 PS Example 4; SEQ ID NO: 32468; 658pp + Sequence Listing; English.  
 PS  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.  
 CC  
 XX  
 SQ Sequence 44 AA;  
 SQ  
 Query Match 31.2%; Score 5; DB 22; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 5; Conservative 0; Mismatches 0; Gaps 0;  
 QY 6 GIDFI 10  
 Db |  
 |  
 |  
 |  
 9 GIDFI 13  
 DE  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Peptide #5905 encoded by probe for measuring cervical gene expression.  
 DE Probe; human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200157278-A2.  
 PN

XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00670.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PA  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI  
 XX WPI; 2001-488901/53.  
 DR  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells -  
 XX  
 PS Claim 27; SEQ ID NO 24297; 487pp; English.  
 PS  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENP; see AAI10068-AA128459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
 CC  
 XX  
 SQ Sequence 44 AA;  
 SQ  
 Query Match 31.2%; Score 5; DB 22; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 GIDFI 10  
 Db |  
 |  
 |  
 |  
 9 GIDFI 13  
 DE  
 DT 17-OCT-2001 (first entry)  
 XX  
 DE Peptide #6459 encoded by probe for measuring placental gene expression.  
 DE Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200157272-A2.  
 PN  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00663.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR



PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488897/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 XX Claim 27; SEQ ID No 32691; 654pp; English.  
 XX The present invention relates to single exon nucleic acid probes (SENP;  
 CC see AAI31315-AA157546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 XX Sequence 44 AA;  
 SQ Query Match 31.2%; Score 5; DB 22; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 GIDFI 10  
 Db 9 GIDFI 13  
 RESULT 9  
 ABG41976  
 ID ABG41976 standard; Peptide; 44 AA.  
 AC ABG41976;  
 DT 19-AUG-2002 (first entry)  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 31641.  
 XX Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX Homo sapiens.  
 OS WO200186003-A2.  
 XX 15-NOV-2001.  
 XX 30-JAN-2001; 2001WO-US00665.  
 XX 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-060840B.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 DR

XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX Claim 27; SEQ ID No 31641; 634pp; English.  
 XX The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence . 44 AA;  
 SQ Query Match 31.2%; Score 5; DB 23; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 GIDFI 10  
 Db 9 GIDFI 13  
 RESULT 10  
 AAB30672  
 ID AAB30672 standard; Protein; 62 AA.  
 XX AAB30672;  
 XX 19-MAR-2001 (first entry)  
 DE Amino acid sequence of a signalling component cassette.  
 XX Membrane-associated protein; transmembrane region; infectious disease;  
 KW HIV infection; inflammatory disease; autoimmunity; rheumatoid arthritis;  
 KW osteoarthritis; inflammatory bowel disease; cancer; allergic disease;  
 KW atopic disease; asthma; eczema; congenital disease; cystic fibrosis;  
 KW sickle cell anaemia; dermatologic disease; psoriasis; neurologic disease;

KW multiple sclerosis; organ transplant rejection; diabetes; CD28;  
 KW graft-versus-host disease.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX WO200063374-A1.  
 PN  
 XX  
 PD 26-OCT-2000.  
 XX  
 PF 17-APR-2000; 2000WO-GB01476.  
 XX  
 PR 16-APR-1999; 99GB-0008816.  
 PR 16-APR-1999; 99GB-0008818.  
 XX  
 PA (CLLT ) CELLTech THERAPEUTICS LTD.  
 XX  
 PI Lawson ADG, Finney HM;  
 XX  
 DR WPI; 2000-687177/67.  
 DR N-PSDB; AAC62383.  
 XX  
 PT Novel nucleic acid useful for treatment or prevention of disease such  
 PT as cancer, infections, asthma, neurological disorder in humans, encodes  
 PT a membrane-associated protein comprising synthetic transmembrane region  
 XX  
 PS Example 1; Fig 2; 42pp; English.  
 XX  
 CC The specification describes nucleic acids encoding a synthetic  
 CC transmembrane region and membrane-associated proteins comprising a  
 CC synthetic transmembrane region. Membrane-associated polypeptides and  
 CC polynucleotides are useful in therapy and for preparing a medicament  
 CC for the treatment or prevention of disease in humans. Diseases or  
 CC disorders which may be treated include infectious diseases e.g. human  
 CC immunodeficiency virus (HIV) infection, inflammatory disease/autoimmunity  
 CC e.g. rheumatoid arthritis, osteoarthritis, inflammatory bowel disease,  
 CC cancer, allergic/atopic disease e.g. asthma, eczema, congenital  
 CC e.g. cystic fibrosis, sickle cell anaemia, dermatologic e.g. psoriasis,  
 CC neurologic e.g. multiple sclerosis, transplants e.g. organ transplant  
 CC rejection, graft-versus-host disease, metabolic/idiopathic disease  
 CC e.g. diabetes. The present sequence represents a signalling component  
 CC cassette, and comprises a human CD28 transmembrane region. The cassette  
 CC is used to facilitate construction of chimeric receptors with modified  
 CC or synthetic transmembrane regions.  
 XX  
 SQ Sequence 62 AA;  
 Query Match 31.2%; Score 5; DB 21; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 FIIWF 13  
 Db 53 FIIWF 57  
 RESULT 11  
 AAB28390  
 ID AAB28390 standard; Protein; 62 AA.  
 XX  
 AC AAB28390;  
 XX  
 DT 19-FEB-2001 (first entry)  
 XX  
 DE Human CD28 transmembrane region fragment.  
 XX  
 KW Oligonucleotide library; sequence block; human;  
 KW CD28 transmembrane region.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200063360-A1.

XX 26-OCT-2000.  
 XX  
 PF 17-APR-2000; 2000WO-GB01498.  
 XX  
 PR 16-APR-1999; 99GB-0008814.  
 XX  
 PA (CLLT ) CELLTech THERAPEUTICS LTD.  
 XX  
 PI Finney HM, Lawson ADG;  
 XX  
 DR WPI; 2000-679595/66.  
 DR N-PSDB; AAC67772.  
 XX  
 PT Generating a library of DNA molecules of varying length and sequence in  
 PT a desired orientation comprises ligating a mixture of double-stranded  
 PT DNA molecules, and cutting ligated molecules in undesired orientations  
 XX  
 PS Example 1; Fig 3; 50pp; English.  
 XX  
 CC The present invention relates to a method for generating a library of  
 CC oligonucleotides of varying length and sequence in a desired orientation.  
 CC The method comprises ligating a double-stranded oligonucleotide mixture  
 CC having 5' and 3' ends compatible to each other and corresponding to  
 CC cleavage products of different but compatible restriction enzymes, and  
 CC cutting ligated oligonucleotide, such that only molecules that are  
 CC ligated in undesired orientations are cut, to produce a sequence block.  
 CC The present sequence is a fragment of human CD28 transmembrane region.  
 CC The coding sequence for the present protein was used to generate a  
 CC cloning cassette, which was used to clone the sequence blocks of the  
 CC present invention.  
 XX  
 SQ Sequence 62 AA;  
 Query Match 31.2%; Score 5; DB 21; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 FIIWF 13  
 Db 53 FIIWF 57  
 RESULT 12  
 AAB84314  
 ID AAB84314 standard; Protein; 62 AA.  
 XX  
 AC AAB84314;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human CD28 protein fragment.  
 XX  
 KW Stimulatory primary signalling motif; immune cell; signal transduction;  
 KW chimeric receptor; inflammatory disease; autoimmune disease; asthma;  
 KW eczema; congenital disease; cystic fibrosis; sickle cell anemia;  
 KW dermatological disease; psoriasis; neurological disease;  
 KW multiple sclerosis; transplant-related disease; metabolic disease;  
 KW organ transplant rejection; graft versus host disease;  
 KW idiopathic disease; diabetes; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200132709-A2.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 01-NOV-2000; 2000WO-GB04183.  
 XX  
 XX 01-NOV-1999; 99GB-0025848.  
 XX  
 PA (CELL-) CELLTech CHIROSCEINCE LTD.

XX PI Finney HM, Lawson ADG;  
 XX DR WPI; 2001-389718/41.  
 XX DR N-PSDB; AAH24845.  
 XX PT Novel cytoplasmic signalling protein and chimeric receptor protein,  
 XX PT useful for treating HIV infection, asthma, eczema, psoriasis, multiple  
 XX PT sclerosis, contain non-natural stimulatory primary signalling motif -  
 XX PS Example 1; Fig 2; 45pp; English.  
 XX CC The present sequence represents a human CD28 extracellular spacer and  
 XX CC transmembrane region. Secondary signalling sequences derived from CD28  
 XX CC can be linked to a non-natural stimulatory primary signalling motif to  
 XX CC produce a cytoplasmic signalling sequence. The primary motif is efficient  
 XX CC at mediating immune cell signal transduction, particularly when  
 XX CC incorporated in an intracellular signalling domain of a chimeric  
 XX CC receptor. The primary signalling motif can be combined in any way so as  
 XX CC to achieve the desired level of activation (or inhibition) of a number  
 XX CC of secondary messenger cascades. The signalling motifs are useful in  
 XX CC therapy and in the manufacture of medicament for treating or preventing  
 XX CC disease in humans or animals. They are useful for treating human  
 XX CC patients suffering from infectious diseases e.g. human immunodeficiency  
 XX CC virus (HIV) infections, inflammatory/autoimmune diseases such as asthma  
 XX CC and eczema, congenital diseases e.g. cystic fibrosis, sickle cell anemia,  
 XX CC dermatological diseases e.g. psoriasis, neurological diseases e.g.  
 XX CC multiple sclerosis, transplant-related disease e.g. organ transplant  
 XX CC rejection, graft versus host disease, metabolic/idiopathic disease  
 XX CC e.g. diabetes, and cancer.  
 XX SQ Sequence 62 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 FIIFF 13  
 Db 53 FIIFF 57  
 RESULT 13  
 AAB98753  
 ID AAB98753 standard; Protein; 62 AA.  
 AC AAB98753;  
 XX DT 07-AUG-2001 (first entry)  
 XX DE Human CD28 extracellular spacer and transmembrane region.  
 XX KW Human; CD28; primary signalling motif; T cell receptor; TCR;  
 KW sequence block; SB; immunosuppressive; secondary signalling sequence;  
 KW antimicrobial; anti-inflammation; dermatological; neuroprotective;  
 KW cytostatic; anti-HIV; antiasthmatic; antiskilling; antipsoriasis;  
 KW antidiabetic; gene therapy; diabetes; immune cell signal transduction;  
 KW infection; inflammation; cancer; autoimmune disease; congenital disease;  
 KW psoriasis; neurological disease; organ transplant rejection.  
 XX OS Homo sapiens.  
 XX PN WO200132867-A1.  
 XX PD 10-MAY-2001.  
 XX PF 01-NOV-2000; 2000WO-GB04193.  
 XX PR 01-NOV-1999; 99GB-0025853.  
 XX PA (CELL-) CELTECH CHIROSCIENCE LTD.  
 XX PI Finney HM, Lawson ADG;

XX WPI; 2001-328791/34.  
 XX DR New nucleic acids encoding polypeptides with expanded primary signaling  
 XX PT motifs, for use in gene therapy, particularly for treating or  
 XX PT preventing infections, inflammations or autoimmune diseases in humans  
 XX PS Example 1; Fig 2; 43pp; English.  
 XX CC The invention relates to novel primary signalling motifs containing  
 XX CC a consensus amino acid sequence. These motifs are extremely  
 XX CC efficient at mediating immune cell signal transduction, particularly  
 XX CC when incorporated into an intracellular signalling domain of a chimeric  
 XX CC receptor. Nucleic acids that encode, and polypeptides that contain,  
 XX CC these primary signalling motifs are useful in medicine and research.  
 XX CC They are useful in therapy, or in the manufacture of a medicament for  
 XX CC treating or preventing disease in humans or in animals. These diseases  
 XX CC include infections (e.g. HIV (human immunodeficiency virus) infection),  
 XX CC inflammatory or autoimmune diseases (e.g. asthma or eczema), congenital  
 XX CC diseases (e.g. cystic fibrosis or sickle cell anaemia), dermatological  
 XX CC diseases (e.g. psoriasis), neurological diseases (e.g. multiple  
 XX CC sclerosis), organ transplant rejection or graft-versus-host disease, or  
 XX CC metabolic/idiopathic diseases (e.g. diabetes or cancer). The  
 XX CC present sequence is the human CD28 extracellular spacer and  
 XX CC transmembrane region used in the construction of a cloning  
 XX CC cassette for generating sequence blocks of primary and secondary  
 XX CC signalling motifs. Primary signalling motifs are sequences that  
 XX CC transduce either a stimulatory or an inhibitory signal, which regulates  
 XX CC primary activation of the T cell receptor (TCR) complex. Secondary  
 XX CC motifs impart secondary or co-stimulatory signalling capacity to a  
 XX CC molecule in T cells.  
 XX SQ Sequence 62 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 FIIFF 13  
 Db 53 FIIFF 57  
 RESULT 14  
 AAB98800  
 ID AAB98800 standard; Protein; 62 AA.  
 XX AC AAB98800;  
 XX DT 08-AUG-2001 (first entry)  
 XX DE Human CD28 extracellular spacer and transmembrane region.  
 XX KW Human; anti-HIV; antiinflammatory; antiasthmatic; dermatological;  
 KW antiskilling; antipsoriasis; neuroprotective; immunosuppressive;  
 KW antidiabetic; cytostatic; anti-HIV infection; inflammation;  
 KW autoimmune disease; cystic fibrosis; sickle cell anaemia; psoriasis;  
 KW neurological disease; organ transplant rejection; diabetes; cancer;  
 KW graft-versus-host disease; organ transplant rejection; diabetes; cancer;  
 KW extracellular spacer; transmembrane region.  
 XX OS Homo sapiens.  
 XX PN WO200132866-A2.  
 XX PD 10-MAY-2001.  
 XX PF 01-NOV-2000; 2000WO-GB04189.  
 XX PR 01-NOV-1999; 99GB-0025854.  
 XX PA (CELL-) CELTECH CHIROSCIENCE LTD.

XX PI Finney HM, Lawson ADG;  
 XX XX WPI; 2001-328790/34.  
 XX XX Novel polynucleotide encoding adaptor receptor protein useful for  
 PT treating human immunodeficiency virus (HIV) infection, asthma, cystic  
 PT fibrosis, multiple sclerosis, organ transplant rejection, diabetes and  
 PT cancer  
 XX  
 PS Example 1; Fig 2; 52pp; English.  
 XX The invention relates to a novel nucleic acid encoding an adaptor  
 CC receptor protein comprising an extracellular ligand-binding domain, a  
 CC transmembrane domain and an intracellular signalling domain. The  
 CC intracellular signalling domain comprises the cytoplasmic portion of at  
 CC least one adaptor protein, and the extracellular ligand-binding domain  
 CC is not CD8 or a major histocompatibility complex (MHC) class I protein.  
 CC The adaptor receptor protein and the nucleic acid encoding it are useful  
 CC in therapy. They are useful in the manufacture of a medicament for the  
 CC treatment or prevention of disease in humans and animals. They are useful  
 CC in the treatment or prevention of infectious diseases (e.g. HIV infection),  
 CC inflammatory and autoimmune diseases (e.g. asthma and eczema),  
 CC congenital diseases (e.g. cystic fibrosis and sickle cell anaemia),  
 CC dermatological diseases (e.g. psoriasis), neurological diseases (e.g.)  
 CC multiple sclerosis), organ transplant rejection, graft-versus-host  
 CC disease and metabolic/idiopathic diseases such as diabetes and cancer.  
 CC The present sequence is the human CD28 extracellular spacer  
 CC and transmembrane region. It was used in the construction of a cloning  
 CC cassette for cloning adaptor molecules of the invention.  
 XX  
 SQ Sequence 62 AA;

Query Match 31.2%; Score 5; DB 22; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY -9 FIIWF 13  
 |||||  
 Db 53 FIIWF 57

RESULT 15  
 AAB28678  
 ID AAB28678 standard; Protein; 62 AA.  
 XX  
 AC AAB28678;  
 XX  
 DT 14-FEB-2001 (first entry)  
 XX  
 DE Protein encoded by cloning cassette system PCR fragment.

XX Membrane-associated protein; antiviral; antibacterial; antiparasitic;  
 KW immunomodulatory; anticancer; antiinflammatory; antiasthmatic;  
 KW antidiabetic; neuroprotective; chimeric receptor; infection;  
 KW inflammatory disorder; autoimmune disorder; cancer; allergy; asthma;  
 KW eczema; cystic fibrosis; sickle cell anaemia; psoriasis;  
 KW multiple sclerosis; organ transplant rejection; diabetes.

XX Homo sapiens.  
 OS Synthetic.

XX WO200063373-A1.

XX 26-OCT-2000.

XX 17-APR-2000; 2000WO-GB01471.

XX 16-APR-1999; 99GB-0008816.

XX (CLLT ) CELLTech THERAPEUTICS LTD.

XX PI Finney HM, Lawson ADG;

XX WPI; 2001-015774/02.  
 DR N-PSDB; AAC65397.  
 XX  
 PT Altering the properties or level of expression of membrane-associated  
 PT proteins, e.g., to change responses to cell surface antigens or the  
 PT sensitivity of intracellular signaling  
 XX  
 PS Example 1; Fig 2; 47pp; English.

XX The nucleotide sequence encoding the present protein was used in the  
 CC construction of chimeric receptors with different binding, extracellular  
 CC spacer, transmembrane and signalling components. The properties and level  
 CC of expression of a membrane-associated protein may be altered by  
 CC substituting a transmembrane region or a membrane-anchoring region for  
 CC transmembrane or membrane-anchoring regions that are not naturally part  
 CC of the protein. The relative response of membrane-associated proteins to  
 CC cell surface-associated antigen versus antigen in solution, and the  
 CC sensitivity of intracellular signaling mediated by membrane-associated  
 CC proteins can be altered. They can be important in treatment of, e.g. HIV  
 CC infection, bacterial infections, parasitic infections,  
 CC inflammatory/autoimmune disorders (e.g. rheumatoid arthritis,  
 CC osteoarthritis or inflammatory bowel disease), cancer, allergic/atopic  
 CC diseases (e.g. asthma or eczema), congenital disorders (e.g. cystic  
 CC fibrosis or sickle cell anaemia), dermatological disorders (e.g. psoriasis),  
 CC neurological disorders (e.g. multiple sclerosis), organ  
 CC transplant rejection, graft-versus-host diseases, or metabolic/idiopathic  
 CC diseases (e.g. diabetes). The polynucleotide encoding the present  
 CC sequence starts with a SpeI site and consists of the extracellular spacer  
 CC h.CD28, the human CD28 transmembrane region, a stop codon, and finishes  
 CC with an EcoRI site.

XX  
 SQ Sequence 62 AA;

Query Match 31.2%; Score 5; DB 22; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIWF 13  
 |||||  
 Db 53 FIIWF 57

RESULT 16  
 AAU98026  
 ID AAU98026 standard; Protein; 62 AA.  
 XX  
 AC AAU98026;  
 XX  
 DT 27-AUG-2002 (first entry)

XX Human IgG1 hinge/CD28 transmembrane domain.

XX Human; IgG1 hinge; CD28; transmembrane domain; CD137; TNFR;  
 KW tumour necrosis factor receptor; cytoplasmic signalling molecule;  
 KW gene therapy; chimeric receptor; human immunodeficiency virus infection;  
 KW HIV; inflammatory disease; autoimmune disease; asthma; eczema; cancer;  
 KW congenital disease; cystic fibrosis; sickle cell anaemia; diabetes;  
 KW dermatological disease; psoriasis; neurological disease;  
 KW multiple sclerosis; transplant rejection; graft-versus-host disease;  
 KW organ transplant rejection; idiopathic disease.

XX Homo sapiens.  
 OS Synthetic.

XX WO200233101-A1.

XX 25-APR-2002.

XX 16-OCT-2001; 2001WO-GB04611.

XX 16-OCT-2000; 2000GB-0025307.

XX PA (CELL-) CELTECH R & D LTD.  
 XX PI  
 XX PI Finney HM, Lawson ADG;  
 XX DR  
 XX DR WPI: 2002-463315/49.  
 XX DR N-PSDB; ABK52921.  
 XX PT Novel nucleic acid encoding cytoplasmic signalling molecule comprising  
 XX PT at least one cytoplasmic signalling sequence derived from Cbl37, useful  
 XX PT for mediating signalling when employed as part of chimaeric receptor  
 XX PT protein  
 XX PS Example 1; Fig 2; 39pp; English.  
 XX CC The invention relates to a nucleic acid (I) encoding a cytoplasmic  
 XX CC signalling molecule comprising at least two cytoplasmic signalling  
 XX CC sequences where at least one cytoplasmic signalling sequence derived from  
 XX CC Cbl37 (a member of the tumour necrosis receptor family, TNFR).  
 XX CC Also included are a nucleic acid (II) encoding a chimaeric receptor  
 XX CC protein which comprises an extracellular ligand-binding domain, a  
 XX CC transmembrane domain and a cytoplasmic signalling domain, where the  
 XX CC cytoplasmic signalling domain is encoded by nucleic acid (I), a vector  
 XX CC comprising a nucleic acid (I) or (II), a host cell containing nucleic  
 XX CC acid (I) or (II) or the vector, a peptide or polypeptide comprising a  
 XX CC cytoplasmic signalling molecule encoded by nucleic acid (I) and a  
 XX CC chimaeric receptor protein encoded by nucleic acid (II). The novel  
 XX CC cytoplasmic signalling molecules encoded by nucleic acid can be used  
 XX CC either by themselves or, as a component part of a larger protein such as  
 XX CC a chimaeric receptor. The cytoplasmic signalling molecules are used  
 XX CC preferentially to mediate signalling when employed as a cytoplasmic  
 XX CC signalling domain of a chimaeric receptor protein. The cytoplasmic  
 XX CC signalling molecules, chimaeric receptors and nucleic acids encoding  
 XX CC them are preferably useful for treating human immunodeficiency virus  
 XX CC (HIV) infection, inflammatory disease/autoimmune disease e.g. asthma, or  
 XX CC eczema, congenital disease e.g. cystic fibrosis, or sickle cell anaemia,  
 XX CC dermatological disease e.g. psoriasis, neurological disease e.g. multiple  
 XX CC sclerosis, transplantation associated diseases e.g. organ transplant  
 XX CC rejection, or graft-versus-host disease, metabolic/idiopathic disease,  
 XX CC e.g. diabetes, or cancer. The present sequence represents a chimaeric  
 XX CC protein consisting of the hinge region of human IgG1 and a region of  
 XX CC human CD28 containing a transmembrane domain. This region is used in the  
 XX CC chimaeric receptor constructs of the invention.  
 XX SQ Sequence 62 AA;  
 Query Match 31.2%; Score 5; DB 23; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 FIIFW 13  
 Db 53 FIIFW 57  
 RESULT 17  
 AAM92938  
 ID AAM92938 standard; Protein; 67 AA.  
 XX AC AAM92938;  
 XX DT 06-NOV-2001 (first entry)  
 XX DE Human digestive system antigen SEQ ID NO: 2287..  
 XX KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
 XX KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
 XX KW digestive system disorder; Meckel's diverticulum.  
 XX OS Homo sapiens.  
 XX PN WO20015314-A2.  
 XX

PD 02-AUG-2001.  
 XX 17-JAN-2001; 2001WO-US01324.  
 PF 31-JAN-2000; 2000US-0179065.  
 XX 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-024617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249219.  
 PR 17-NOV-2000; 2000US-0249220.  
 PR 17-NOV-2000; 2000US-0249221.  
 PR 17-NOV-2000; 2000US-0249222.  
 PR 17-NOV-2000; 2000US-0249223.  
 PR 17-NOV-2000; 2000US-0249224.  
 PR 17-NOV-2000; 2000US-0249225.  
 PR 17-NOV-2000; 2000US-0249226.  
 PR 17-NOV-2000; 2000US-0249227.  
 PR 17-NOV-2000; 2000US-0249228.  
 PR 17-NOV-2000; 2000US-0249229.  
 PR 17-NOV-2000; 2000US-0249230.  
 PR 17-NOV-2000; 2000US-0249231.  
 PR 17-NOV-2000; 2000US-0249232.  
 PR 17-NOV-2000; 2000US-0249233.  
 PR 17-NOV-2000; 2000US-0249234.  
 PR 17-NOV-2000; 2000US-0249235.  
 PR 17-NOV-2000; 2000US-0249236.  
 PR 17-NOV-2000; 2000US-0249237.  
 PR 17-NOV-2000; 2000US-0249238.  
 PR 17-NOV-2000; 2000US-0249239.  
 PR 17-NOV-2000; 2000US-0249240.  
 PR 17-NOV-2000; 2000US-0249241.  
 PR 17-NOV-2000; 2000US-0249242.  
 PR 17-NOV-2000; 2000US-0249243.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249246.  
 PR 17-NOV-2000; 2000US-0249247.  
 PR 17-NOV-2000; 2000US-0249248.  
 PR 17-NOV-2000; 2000US-0249249.  
 PR 17-NOV-2000; 2000US-0249250.  
 PR 17-NOV-2000; 2000US-0249251.  
 PR 17-NOV-2000; 2000US-0249252.  
 PR 17-NOV-2000; 2000US-0249253.  
 PR 17-NOV-2000; 2000US-0249254.  
 PR 17-NOV-2000; 2000US-0249255.  
 PR 17-NOV-2000; 2000US-0249256.  
 PR 17-NOV-2000; 2000US-0249257.  
 PR 17-NOV-2000; 2000US-0249258.  
 PR 17-NOV-2000; 2000US-0249259.  
 PR 17-NOV-2000; 2000US-0249260.  
 PR 17-NOV-2000; 2000US-0249261.  
 PR 17-NOV-2000; 2000US-0249262.  
 PR 17-NOV-2000; 2000US-0249263.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249266.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249268.  
 PR 17-NOV-2000; 2000US-0249269.  
 PR 17-NOV-2000; 2000US-0249270.  
 PR 17-NOV-2000; 2000US-0249271.  
 PR 17-NOV-2000; 2000US-0249272.  
 PR 17-NOV-2000; 2000US-0249273.  
 PR 17-NOV-2000; 2000US-0249274.  
 PR 17-NOV-2000; 2000US-0249275.  
 PR 17-NOV-2000; 2000US-0249276.  
 PR 17-NOV-2000; 2000US-0249277.  
 PR 17-NOV-2000; 2000US-0249278.  
 PR 17-NOV-2000; 2000US-0249279.  
 PR 17-NOV-2000; 2000US-0249280.  
 PR 17-NOV-2000; 2000US-0249281.  
 PR 17-NOV-2000; 2000US-0249282.  
 PR 17-NOV-2000; 2000US-0249283.  
 PR 17-NOV-2000; 2000US-0249284.  
 PR 17-NOV-2000; 2000US-0249285.  
 PR 17-NOV-2000; 2000US-0249286.  
 PR 17-NOV-2000; 2000US-0249287.  
 PR 17-NOV-2000; 2000US-0249288.  
 PR 17-NOV-2000; 2000US-0249289.  
 PR 17-NOV-2000; 2000US-0249290.  
 PR 17-NOV-2000; 2000US-0249291.  
 PR 17-NOV-2000; 2000US-0249292.  
 PR 17-NOV-2000; 2000US-0249293.  
 PR 17-NOV-2000; 2000US-0249294.  
 PR 17-NOV-2000; 2000US-0249295.  
 PR 17-NOV-2000; 2000US-0249296.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249298.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0251989.  
 PR 06-DEC-2000; 2000US-0256719.  
 PR 08-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251899.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0255978.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-502630/55.

DR N-PSDB; AAK88711.  
 XX  
 PT Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognosing disorders of the  
 PT digestive system, particularly cancer and cancer metastases -  
 XX  
 PS Claim 11; SEQ ID NO 2287; 986pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer, Meckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a digestive system antigen of  
 CC the invention.  
 XX  
 SQ Sequence 67 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 67;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 DFIIF 12  
 Db 32 DFIIF 36  
 RESULT 18  
 AAU20034  
 ID AAU20034 standard; Protein; 67 AA.  
 XX AC AAU20034;  
 XX  
 DT 06-DEC-2001 (first entry)  
 XX  
 DE Human liver associated polypeptide #65.  
 XX  
 KW Liver associated protein; human; mouse; rabbit; goat; horse; cat; dog;  
 KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;  
 KW antihemetic; antiproliferative; cytostatic; cardiant; neuroprotective;  
 KW cerebroprotective; nontropic; antibacterial; virucide; fungicide; cancer;  
 KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;  
 KW hyperproliferative disorder; breast; liver; cardiovascular disorder;  
 KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
 KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
 KW gastrointestinal disorder; renal disorder; respiratory disorder;  
 KW wound healing; skin aging; organ transplantation; tissue regeneration;  
 KW anti-infertility.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155355-A1.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01351.  
 XX  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.

26-JUL-2000;	2000US-02209663;	PR
26-JUL-2000;	2000US-02209664;	PR
14-AUG-2000;	2000US-02245118;	PR
14-AUG-2000;	2000US-02245119;	PR
14-AUG-2000;	2000US-02252113;	PR
14-AUG-2000;	2000US-02252114;	PR
14-AUG-2000;	2000US-02252166;	PR
14-AUG-2000;	2000US-02252667;	PR
14-AUG-2000;	2000US-02252668;	PR
14-AUG-2000;	2000US-02252760;	PR
14-AUG-2000;	2000US-02252761;	PR
14-AUG-2000;	2000US-02254477;	PR
14-AUG-2000;	2000US-02254757;	PR
14-AUG-2000;	2000US-02257558;	PR
14-AUG-2000;	2000US-02257559;	PR
18-AUG-2000;	2000US-02262799;	PR
22-AUG-2000;	2000US-02266681;	PR
22-AUG-2000;	2000US-02266688;	PR
22-AUG-2000;	2000US-02271182;	PR
23-AUG-2000;	2000US-02270009;	PR
23-AUG-2000;	2000US-02289824;	PR
01-SEP-2000;	2000US-02292877;	PR
01-SEP-2000;	2000US-02293433;	PR
01-SEP-2000;	2000US-02293444;	PR
01-SEP-2000;	2000US-02293445;	PR
05-SEP-2000;	2000US-02295059;	PR
05-SEP-2000;	2000US-02295113;	PR
06-SEP-2000;	2000US-02304337;	PR
06-SEP-2000;	2000US-02304338;	PR
08-SEP-2000;	2000US-02311242;	PR
08-SEP-2000;	2000US-02311243;	PR
08-SEP-2000;	2000US-02311244;	PR
08-SEP-2000;	2000US-02311413;	PR
08-SEP-2000;	2000US-02311414;	PR
08-SEP-2000;	2000US-02320880;	PR
08-SEP-2000;	2000US-02320881;	PR
12-SEP-2000;	2000US-02321968;	PR
14-SEP-2000;	2000US-02323997;	PR
14-SEP-2000;	2000US-02323998;	PR
14-SEP-2000;	2000US-02323999;	PR
14-SEP-2000;	2000US-02324000;	PR
14-SEP-2000;	2000US-02324001;	PR
14-SEP-2000;	2000US-02324003;	PR
14-SEP-2000;	2000US-02330683;	PR
14-SEP-2000;	2000US-02330684;	PR
21-SEP-2000;	2000US-02330685;	PR
21-SEP-2000;	2000US-02342223;	PR
21-SEP-2000;	2000US-02342274;	PR
23-SEP-2000;	2000US-02343997;	PR
23-SEP-2000;	2000US-02343998;	PR
26-SEP-2000;	2000US-02354884;	PR
27-SEP-2000;	2000US-02358334;	PR
27-SEP-2000;	2000US-02358336;	PR
29-SEP-2000;	2000US-02363277;	PR
29-SEP-2000;	2000US-02363667;	PR
29-SEP-2000;	2000US-02363668;	PR
29-SEP-2000;	2000US-02363669;	PR
29-SEP-2000;	2000US-02363700;	PR
02-OCT-2000;	2000US-02368602;	PR
02-OCT-2000;	2000US-02370337;	PR
02-OCT-2000;	2000US-02370338;	PR
02-OCT-2000;	2000US-02370339;	PR
02-OCT-2000;	2000US-02370400;	PR
13-OCT-2000;	2000US-02393935;	PR
13-OCT-2000;	2000US-02393937;	PR
20-OCT-2000;	2000US-02409960;	PR
20-OCT-2000;	2000US-02411221;	PR
20-OCT-2000;	2000US-02411785;	PR
20-OCT-2000;	2000US-02417866;	PR
20-OCT-2000;	2000US-02417877;	PR
20-OCT-2000;	2000US-02418008;	PR
20-OCT-2000;	2000US-02418009;	PR
20-OCT-2000;	2000US-02418826;	PR
01-NOV-2000;	2000US-02446177;	PR
08-NOV-2000;	2000US-02464774;	PR

PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	03-DEC-2000;	2000US-0251030.
PR	03-DEC-2000;	2000US-0251988.
PR	06-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC

33

CC disorders such as premature labour and infertility, gastrointestinal  
 CC disorders such as Crohn's disease, renal disorders, such as  
 CC glomerulonephritis and respiratory disorders such as asthma and pleurisy.  
 CC The polypeptides can also be used to aid wound healing, to prevent skin  
 CC aging due to sunburn, to maintain organs before transplantation, to  
 CC regenerate tissues and in chemotaxis.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 31.2%; Score 5; DB 22; Length 67;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 DFIIF 12  
 Db 32 DFIIF 36  
 RESULT 19  
 ABP40895  
 ID ABP40895 standard; Protein; 67 AA.  
 XX  
 AC ABP40895;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Human liver antigen HLIOL82, SEQ ID NO:221.  
 XX  
 KW Human; liver antigen; liver disorder; hepatic disorder; infection;  
 KW hepatitis; viral; parasitic; bacterial; fungal; inflammatory condition;  
 KW cirrhosis; granulomatous hepatitis; toxin damage; drug damage;  
 KW autoimmune disease; Wilson's disease; primary biliary cirrhosis;  
 KW neoplastic disorder; cancer; tumour; portal hypertension;  
 KW gastrointestinal disorder; hepatitis; drug screening; gene therapy;  
 KW chromosome mapping; forensic analysis; antibody preparation;  
 KW hepatotropic; cytotoxic; antiinflammatory; virucide; antibacterial;  
 KW fungicide; parasiticide; antidote; immunosuppressive.  
 OS Homo sapiens.

XX  
 XX  
 PN US2002042096-A1.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 17-JAN-2001; 2001US-0764887.  
 XX  
 PP 31-JAN-2000; 2000US-179065P.  
 PR 04-FEB-2000; 2000US-180628P.  
 PR 28-JUN-2000; 2000US-214886P.  
 PR 07-JUL-2000; 2000US-216647P.  
 PR 07-JUL-2000; 2000US-216880P.  
 PR 11-JUL-2000; 2000US-217487P.  
 PR 11-JUL-2000; 2000US-217496P.  
 PR 14-JUL-2000; 2000US-218290P.  
 PR 26-JUL-2000; 2000US-220963P.  
 PR 14-AUG-2000; 2000US-220964P.  
 PR 14-AUG-2000; 2000US-224518P.  
 PR 14-AUG-2000; 2000US-224519P.  
 PR 14-AUG-2000; 2000US-225267P.  
 PR 14-AUG-2000; 2000US-225268P.  
 PR 14-AUG-2000; 2000US-225270P.  
 PR 14-AUG-2000; 2000US-225447P.  
 PR 14-AUG-2000; 2000US-225757P.  
 PR 14-AUG-2000; 2000US-225758P.  
 PR 22-AUG-2000; 2000US-226868P.  
 PR 30-AUG-2000; 2000US-228924P.  
 PR 01-SEP-2000; 2000US-229287P.  
 PR 01-SEP-2000; 2000US-229343P.  
 PR 01-SEP-2000; 2000US-229344P.  
 PR 01-SEP-2000; 2000US-229345P.  
 PR 05-SEP-2000; 2000US-229509P.

PR 05-SEP-2000; 2000US-229513P.  
 PR 08-SEP-2000; 2000US-231413P.  
 PR 21-SEP-2000; 2000US-234223P.  
 PR 21-SEP-2000; 2000US-234274P.  
 PR 25-SEP-2000; 2000US-234997P.  
 PR 27-SEP-2000; 2000US-235834P.  
 PR 29-SEP-2000; 2000US-236327P.  
 PR 29-SEP-2000; 2000US-236367P.  
 PR 29-SEP-2000; 2000US-236368P.  
 PR 29-SEP-2000; 2000US-236369P.  
 PR 02-OCT-2000; 2000US-236370P.  
 PR 02-OCT-2000; 2000US-236802P.  
 PR 02-OCT-2000; 2000US-237037P.  
 PR 02-OCT-2000; 2000US-237038P.  
 PR 02-OCT-2000; 2000US-237039P.  
 PR 13-OCT-2000; 2000US-237040P.  
 PR 20-OCT-2000; 2000US-239935P.  
 PR 20-OCT-2000; 2000US-240960P.  
 PR 20-OCT-2000; 2000US-241785P.  
 PR 20-OCT-2000; 2000US-241809P.  
 PR 17-NOV-2000; 2000US-244617P.  
 PR 08-DEC-2000; 2000US-249299P.  
 PR 08-DEC-2000; 2000US-251856P.  
 PR 08-DEC-2000; 2000US-251868P.  
 PR 08-DEC-2000; 2000US-251869P.  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 DR WPI; 2002-381944/41.  
 DR N-PSDB; ABN90100.  
 XX  
 DR New nucleic acid encoding human liver antigens, useful for diagnosis,  
 XX treatment and prevention of e.g. hepatitis and hepatic cancer, also  
 XX related polypeptides and antibodies

Claim 11; SEQ ID NO 221; 181pp; English.

XX The invention relates to 145 novel human liver antigens (ABP40831-  
 CC ABP40975) and to cDNAs encoding them (ABN90036-ABN90180), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human liver antigen  
 CC polynucleotides, antibodies against human liver antigens, and the use of  
 CC liver antigen polynucleotides and polypeptides in diagnosing, treating,  
 CC prognosing or preventing various disorders of the liver. Such conditions  
 CC include viral infections (e.g., cytomegalovirus, Epstein-Barr virus,  
 CC hepatitis A virus, hepatitis B virus and hepatitis C virus), parasitic  
 CC infections (e.g., Clonorchis sinensis, Echinococcus granulosus and  
 CC Entamoeba histolytica), and also bacterial and fungal infections. Other  
 CC disorders that may be treated include inflammatory conditions (e.g.,  
 CC autoimmune diseases (e.g., adenomas, haemangiomas and hepatocellular  
 CC carcinomas), portal hypertension, or gastrointestinal disorders (e.g.,  
 CC peptic ulcers, gastritis and peritoneal diseases). Liver antigen  
 CC polypeptides and polynucleotides may also be used in screening for  
 CC compounds which modulate liver antigen expression or activity. The  
 CC polynucleotides may further be used for gene therapy, chromosome  
 CC mapping, in the identification of individuals and in forensic analysis,  
 CC and the polypeptides may be used as molecular weight markers or to  
 CC prepare antibodies useful in disease diagnosis, drug targeting and  
 CC phenotyping. The present sequence represents a human liver antigen of  
 CC the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO at seqdata.uspto.gov/sequence/  
 XX

Sequence 67 AA;  
 SQ



Query Match 31.2%; Score 5; DB 23; Length 67;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
Db 32 DFIIF 36

RESULT 20  
AAB58985  
ID AAB58985 standard; Protein; 68 AA.  
XX AC AAB58985;  
XX  
DT 27-MAR-2001 (first entry)  
XX  
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 693.  
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antiinflammatory; antiulcer; vulnary; anticonvulsant;  
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;  
KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
KW cardiovascular disorder; wound healing; neurological disease.

OS Homo sapiens.  
PN WO200055173-A1.  
XX

PD 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05881.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-611515/58.

XX N-PSDB; AAF21888.

XX New human breast and ovarian cancer associated gene sequences and the  
PT polypeptides encoded by these genes, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -

XX Claim 11; Page 1147; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
CC associated with breast and ovarian cancer. Included in the invention are  
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
CC isolation and characterisation of the DNA and protein sequences of the  
CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
CC antidiabetic; antiinflammatory; antiulcer; vulnary; anticonvulsant;  
CC antibacterial; antifungal; antiparasitic and cardiac activity. The  
CC polynucleotide and protein sequences are used in the diagnosis of  
CC particularly breast and ovarian cancer. The nucleic acid sequences,  
CC proteins, agonists and antagonists may also be used in the diagnosis,  
CC prevention and treatment of immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemia; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases.

XX Sequence 68 AA;

Query Match 31.2%; Score 5; DB 21; Length 68;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
Db 47 DFIIF 51

RESULT 21  
AAR94651  
ID AAR94651 standard; Protein; 69 AA.  
XX AC AAR94651;  
XX  
DT 08-DEC-1996 (first entry)  
XX  
DE Mouse Fas-associated protein PTP-BAS type 5b (MFAP23).  
XX Fas-associated protein; tumour necrosis factor receptor; PTP-BAS;  
KW apoptosis; FAP; cell surface protein; autoimmune disease; HIV virus;  
KW hybridoma; cell death.  
OS Mus musculus.  
XX  
PH Key Location/Qualifiers  
FT Misc-difference 34..69 /note= "AAS diverging from PTP-BAS family"  
FT

XX WO9534661-A1.

XX 21-DEC-1995.

XX 14-JUN-1995; 95WO-US07583.

XX 27-MAR-1995; 95US-0410804.

XX 14-JUN-1994; 94US-0259514.

XX (LJOL-) LA JOLLA CANCER RES FOUND.

XX Reed JC, Sato T;

XX WPI; 1996-049689/05.

XX N-PSDB; AAT18385.

XX New Fas associated proteins PTP-BAS types 4 and 5 - involved in  
PT programmed cell death, used for modulating apoptosis, e.g. for  
PT treating cancer, and for identifying other modulators

XX Claim 17; Fig.16; 84pp; English.

XX Regulation of apoptosis can be used in the treatment of cancer,  
CC autoimmune disease and viral disease (e.g. HIV virus infections),  
CC also to prolong survival of e.g. hybridoma cells in culture  
CC resulting in an increase in bioproduct yield. This protein may  
CC also be used in a method for diagnosing diseases associated with  
CC increased or decreased levels of Fas-associated protein.

XX Sequence 69 AA;

Query Match 31.2%; Score 5; DB 17; Length 69;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFIIF 11  
Db 58 IDFIIF 62

RESULT 22  
AAG00588  
ID AAG00588 standard; Protein; 72 AA.

```

XX AC AAG00588;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein, SEQ ID NO: 4669.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PS (GEST) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.
XX DR N-PSDB; AAC00594.
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 13; SEQ ID 4669; 71pp + CD-ROM; English.
XX CC The present sequence is a polypeptide encoded by one of a large number
XX CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
XX CC were prepared from total human RNAs or polyA+ RNAs derived from 30
XX CC different tissues. EST sequences usually correspond mainly to the 3'
XX CC untranslated region (UTR) of the mRNA because they are often obtained
XX CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
XX CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
XX CC those cases where longer cDNA sequences have been obtained, the full 5'
XX CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
XX CC ends and can therefore be used to obtain full length cDNAs and genomic
XX CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
XX CC chromosome mapping procedures. They are used to obtain upstream
XX CC regulatory sequences and to design expression and secretion vectors.
XX SQ Sequence 72 AA;
XX CC Query Match 31.2%; Score 5; DB 21; Length 72;
XX CC Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX CC Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 IDFTI 11
DB 12 IDFTI 16
XX DE
XX DE RESULT 23
XX DE ABP40821
XX ID ABP40821 standard; Protein; 74 AA.
XX AC ABP40821;
XX DT 24-JUL-2002 (first entry)
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5666.
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX KW antibacterial; gene therapy.
XX OS Staphylococcus epidermidis.
XX PI
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.

```

```

PN US6380370-B1.
XX XX
XX PD 30-APR-2002.
XX PF 13-AUG-1998; 98US-0134001.
XX PR 14-AUG-1997; 97US-055779P.
XX PR 08-NOV-1997; 97US-064964P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm LA, Bush D;
XX DR WPI; 2002-381255/41.
XX DR N-FSDB; ABN93366.
XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX PS polypeptide, useful for diagnosing and treating bacterial infections -
XX CC Disclosure; SEQ ID 5666; 267pp; English.
XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX CC antibacterial activity and can be used in gene therapy. The sequences
XX CC can also be used in the diagnosis and treatment of bacterial infections,
XX CC particularly S. epidermidis infections. The sequences can be used to
XX CC screen for compounds able to interfere with the S. epidermidis life
XX CC cycle or inhibit S. epidermidis infection.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC USPTO web site.
XX SQ Sequence 74 AA;
XX CC Query Match 31.2%; Score 5; DB 23; Length 74;
XX CC Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX CC Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 FIIFW 13
DB 40 FIIFW 44
XX DE
XX DE RESULT 24
XX DE ABG27612
XX ID ABG27612 standard; Protein; 76 AA.
XX AC ABG27612;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #27603.
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.

```

DR N-PSDB; AAS91799.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 20; SEQ ID No 57971; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABC00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 76 AA;

QY Query Match 31.2%; Score 5; DB 22; Length 76;

Db Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10

Db 9 GIDFI 13

RESULT 25

AAW28161

ID AAW28161 standard; Protein: 81 AA.

AC AAW28161;

XX 27-AUG-1998 (first entry)

DE Amino acid sequence of a large conductance mechanosensitive channel.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;

KW Staphylococcal gene; regulatory element; bacterial gene expression;

KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;

KW toxic shock syndrome.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT Misc-difference 1..81

FT /note= "residues designated X are not defined in

FT the specification"

XX WO9730070-A1.

PN 21-AUG-1997.

PD 19-FEB-1997; 97WO-US02318.

PF 20-FEB-1996; 96US-0011888.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PA

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;

PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI: 1997-424969/39.

DR N-PSDB; AAT84086.

XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used

PT to isolate antimicrobial compounds, and in vaccines against S.

PT aureus infection

XX Claim 6; Page 504; 989pp; English.

PS The present sequence represents a Staphylococcus aureus protein, that,

XX based on homology with an Escherichia coli protein, is believed to be a

CC large conductance mechanosensitive channel. The DNA sequence was isolated

CC from a library of clones of S. aureus WCUH 29 in Escherichia coli. The

CC DNA sequence can be used in the construction of ribozymes and antisense

CC sequences to control the expression of regulatory elements for the

CC sequence is also useful as a source of regulatory elements for the

CC control of bacterial gene expression. The present protein may be used

CC to produce vaccines to enable a host to produce specific antibodies

CC with antibacterial action. These vaccines and antibodies would protect

CC a host against invasion by S. aureus, and conditions relating to

CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled

CC skin syndrome, and toxic shock syndrome.

XX Sequence 81 AA;

QY Query Match 31.2%; Score 5; DB 18; Length 81;

Db Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11

Db 32 IDFI 36

RESULT 26

AAW94589

ID AAW94589 standard; Protein: 82 AA.

AC AAW94589;

XX 21-NOV-2001 (first entry)

DE Human reproductive system related antigen SEQ ID NO: 3247.

XX Human; reproductive system related antigen; reproductive system disorder;

KW cancer; gene therapy.

XX Homo sapiens.

XX WO200155320-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0225759.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-02331968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 25-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 02-OCT-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-024475.  
PR 08-NOV-2000; 2000US-024476.  
PR 08-NOV-2000; 2000US-024477.  
PR 08-NOV-2000; 2000US-024478.  
PR 08-NOV-2000; 2000US-024523.  
PR 08-NOV-2000; 2000US-024524.  
PR 08-NOV-2000; 2000US-024525.  
PR 08-NOV-2000; 2000US-024526.  
PR 08-NOV-2000; 2000US-024527.  
PR 08-NOV-2000; 2000US-024528.  
PR 08-NOV-2000; 2000US-024532.  
PR 08-NOV-2000; 2000US-0245609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0254097.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465570/50.  
XX N-PSDB; AAL00559.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen  
XX is used in preventing, treating or ameliorating a medical condition.  
XX  
XX Claim 11; SEQ ID NO 3247; 1297pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders,  
XX including cancer. The present sequence is a protein of the invention.  
XX  
XX Sequence 82 AA;  
XX

Query Match 31.2%; Score 5; DB 22; Length 82;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





PI Tam RC;  
XX WPI; 1999-443609/37.  
DR N-PSDB; AAX90349.  
XX Treatment of immune system-mediated diseases by inhibiting  
PT expression of CD28, IL-2, gamma-interferon or IL-8  
XX  
XX Disclosure; Fig 1C; 45pp; English.  
XX  
XX The present invention describes a method for inhibiting the expression  
CC of CD28, IL-2, gamma-interferon or IL-8 in a mammal. The method  
CC comprises subcutaneous administration of an oligonucleotide (OGN).  
CC AAX90288 to AAX90291 represent specifically claimed OGNs for use in the  
CC method. The OGNs are used for the treatment of immune system-mediated  
CC diseases. The present sequence represents a protein sequence of the  
CC human CD28 gene from the present invention.  
XX  
XX Sequence 103 AA;  
SQ  
Query Match 31.2%; Score 5; DB 20; Length 103;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 FIIWF 13  
Db 57 FIIWF 61  
RESULT 31  
AAM94419  
ID AAM94419 standard; Protein; 104 AA.  
XX  
XX AAM94419;  
AC  
XX  
XX 21-NOV-2001 (first entry)  
DT  
XX  
XX Human reproductive system related antigen SEQ ID NO: 3077.  
XX  
XX Human reproductive system related antigen;  
KW cancer; gene therapy.  
XX  
XX Homo sapiens.  
OS  
XX WO200155320-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01339.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR  
XX 04-FEB-2000; 2000US-0180628.  
PR  
XX 24-FEB-2000; 2000US-0184664.  
PR  
XX 02-MAR-2000; 2000US-0186350.  
PR  
XX 16-MAR-2000; 2000US-0189874.  
PR  
XX 17-MAR-2000; 2000US-0190076.  
PR  
XX 18-APR-2000; 2000US-0198123.  
PR  
XX 19-MAY-2000; 2000US-0205515.  
PR  
XX 07-JUN-2000; 2000US-0209467.  
PR  
XX 28-JUN-2000; 2000US-0214886.  
PR  
XX 30-JUN-2000; 2000US-0215135.  
PR  
XX 07-JUL-2000; 2000US-0216647.  
PR  
XX 07-JUL-2000; 2000US-0216880.  
PR  
XX 11-JUL-2000; 2000US-0217487.  
PR  
XX 11-JUL-2000; 2000US-0217496.  
PR  
XX 14-JUL-2000; 2000US-0218290.  
PR  
XX 26-JUL-2000; 2000US-0220963.  
PR  
XX 26-JUL-2000; 2000US-0220964.  
PR  
XX 14-AUG-2000; 2000US-0224518.  
PR  
XX 14-AUG-2000; 2000US-0224519.  
PR  
XX 14-AUG-2000; 2000US-0225213.  
PR  
XX 14-AUG-2000; 2000US-0225214.  
PR  
XX 14-AUG-2000; 2000US-0225266.

PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX  
XX WPI; 2001-483426/52.  
DR N-PSDB; AAK63524.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
XX Claim 11; SEQ ID NO 18336; 3071pp + Sequence Listing; English.  
PS  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX  
XX Sequence 92 AA;  
SQ  
Query Match 31.2%; Score 5; DB 22; Length 92;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 ANCGI 7  
Db 18 ANCGI 22  
RESULT 30  
AAY24470  
ID AAY24470 standard; Protein; 103 AA.  
XX  
XX AAY24470;  
AC  
XX  
XX 24-SEP-1999 (first entry)  
DT  
XX  
XX Human CD28 gene protein sequence #2.  
DE  
XX  
XX CD28; inhibition; antisense oligonucleotide; interleukin 2; IL-2;  
KW immune system mediated disease; gamma-interferon; IL-8.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX US5932556-A.  
PN  
XX  
XX 03-AUG-1999.  
PD  
XX  
XX 18-SEP-1995; 95US-0529878.  
PF  
XX  
XX 18-SEP-1995; 95US-0529878.  
PR  
XX  
XX (TAMR/) TAM R C.  
PA  
XX

PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232297.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 26-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 02-OCT-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2000US-0254097.  
PR 05-JAN-2001; 2001US-02559678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

N-PSDB; AAL00389.

Isolated nucleic acid molecule encoding a reproductive system antigen -  
is used in preventing, treating or ameliorating a medical condition -

Claim 11; SEQ ID NO 3077; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a  
number of human reproductive system related antigens. These can be used  
in the prevention and treatment of reproductive system disorders,  
including cancer. The present sequence is a protein of the invention.

Sequence 104 AA;

Query Match 31.2%; Score 5; DB 22; Length 104;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11

Db 31 IDFII 35

RESULT 32

ABG30130

ID ABG30130 standard; Protein; 111 AA.

XX



us-09-854-133-587.olig.rag

Tue, May 13 12:12:54 2003

AC ABG30130;  
 XX 18-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #30121.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI: 2001-639362/73.  
 XX N-PSDB; AAS94317.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 60489; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 111 AA;  
 SQ  
 Query Match 31.2%; Score 5; DB 22; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 3 ANCGI 7  
 Db 104 ANCGI 108  
 RESULT 33  
 AAG40079  
 ID AAG40079 standard; Protein; 112 AA.  
 XX  
 AC AAG40079;  
 XX

18-OCT-2000 (first entry)  
 Arabidopsis thaliana protein fragment SEQ ID NO: 49681.  
 Protein identification; signal transduction pathway; metabolic pathway;  
 hybridisation assay; genetic mapping; gene expression control; promoter;  
 termination sequence.  
 Arabidopsis thaliana.  
 Arabidopsis thaliana.  
 EP1033405-A2.  
 06-SEP-2000.  
 25-FEB-2000; 2000EP-0301439.  
 25-FEB-1999; 99US-0121825.  
 05-MAR-1999; 99US-0123180.  
 09-MAR-1999; 99US-0123548.  
 23-MAR-1999; 99US-0125788.  
 25-MAR-1999; 99US-0126264.  
 29-MAR-1999; 99US-0126785.  
 01-APR-1999; 99US-0127462.  
 06-APR-1999; 99US-0128234.  
 08-APR-1999; 99US-0128714.  
 16-APR-1999; 99US-0129845.  
 19-APR-1999; 99US-0130077.  
 21-APR-1999; 99US-0130449.  
 23-APR-1999; 99US-0130510.  
 28-APR-1999; 99US-0130891.  
 30-APR-1999; 99US-0131449.  
 30-APR-1999; 99US-0132048.  
 30-APR-1999; 99US-0132407.  
 04-MAY-1999; 99US-0132484.  
 05-MAY-1999; 99US-0132485.  
 06-MAY-1999; 99US-0132486.  
 07-MAY-1999; 99US-0132487.  
 11-MAY-1999; 99US-0132863.  
 14-MAY-1999; 99US-0134256.  
 14-MAY-1999; 99US-0134219.  
 14-MAY-1999; 99US-0134219.  
 14-MAY-1999; 99US-0134221.  
 14-MAY-1999; 99US-0134370.  
 18-MAY-1999; 99US-0134768.  
 19-MAY-1999; 99US-0134941.  
 20-MAY-1999; 99US-0135124.  
 21-MAY-1999; 99US-0135353.  
 24-MAY-1999; 99US-0135629.  
 25-MAY-1999; 99US-0136021.  
 27-MAY-1999; 99US-0136392.  
 28-MAY-1999; 99US-0136782.  
 01-JUN-1999; 99US-0137222.  
 03-JUN-1999; 99US-0137528.  
 04-JUN-1999; 99US-0137502.  
 07-JUN-1999; 99US-0137724.  
 08-JUN-1999; 99US-0138094.  
 10-JUN-1999; 99US-0138540.  
 10-JUN-1999; 99US-0138847.  
 14-JUN-1999; 99US-0139119.  
 16-JUN-1999; 99US-0139452.  
 16-JUN-1999; 99US-0139453.  
 17-JUN-1999; 99US-0139492.  
 18-JUN-1999; 99US-0139454.  
 18-JUN-1999; 99US-0139455.  
 18-JUN-1999; 99US-0139456.  
 18-JUN-1999; 99US-0139457.  
 18-JUN-1999; 99US-0139458.  
 18-JUN-1999; 99US-0139459.  
 18-JUN-1999; 99US-0139460.  
 18-JUN-1999; 99US-0139461.  
 18-JUN-1999; 99US-0139462.  
 18-JUN-1999; 99US-0139463.  
 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 24-JUN-1999; 99US-0140354.  
 PR 25-JUN-1999; 99US-0140695.  
 PR 26-JUN-1999; 99US-0140823.  
 PR 27-JUN-1999; 99US-0140991.  
 PR 28-JUN-1999; 99US-0141287.  
 PR 29-JUN-1999; 99US-0141842.  
 PR 30-JUN-1999; 99US-0142055.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142390.  
 PR 03-JUL-1999; 99US-0142803.  
 PR 04-JUL-1999; 99US-0142920.  
 PR 05-JUL-1999; 99US-0142977.  
 PR 06-JUL-1999; 99US-0143542.  
 PR 07-JUL-1999; 99US-0143624.  
 PR 08-JUL-1999; 99US-0144005.  
 PR 09-JUL-1999; 99US-0144085.  
 PR 10-JUL-1999; 99US-0144086.  
 PR 11-JUL-1999; 99US-0144325.  
 PR 12-JUL-1999; 99US-0144331.  
 PR 13-JUL-1999; 99US-0144332.  
 PR 14-JUL-1999; 99US-0144333.  
 PR 15-JUL-1999; 99US-0144334.  
 PR 16-JUL-1999; 99US-0144335.  
 PR 17-JUL-1999; 99US-0144352.  
 PR 18-JUL-1999; 99US-0144632.  
 PR 19-JUL-1999; 99US-0144884.  
 PR 20-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 22-JUL-1999; 99US-0145088.  
 PR 23-JUL-1999; 99US-0145085.  
 PR 24-JUL-1999; 99US-0145087.  
 PR 25-JUL-1999; 99US-0145089.  
 PR 26-JUL-1999; 99US-0145192.  
 PR 27-JUL-1999; 99US-0145192.  
 PR 28-JUL-1999; 99US-0145218.  
 PR 29-JUL-1999; 99US-0145224.  
 PR 30-JUL-1999; 99US-0145276.  
 PR 31-JUL-1999; 99US-0145913.  
 PR 01-AUG-1999; 99US-0145918.  
 PR 02-AUG-1999; 99US-0145919.  
 PR 03-AUG-1999; 99US-0145951.  
 PR 04-AUG-1999; 99US-0146386.  
 PR 05-AUG-1999; 99US-0146388.  
 PR 06-AUG-1999; 99US-0146389.  
 PR 07-AUG-1999; 99US-0147038.  
 PR 08-AUG-1999; 99US-0147204.  
 PR 09-AUG-1999; 99US-0147302.  
 PR 10-AUG-1999; 99US-0147192.  
 PR 11-AUG-1999; 99US-0147260.  
 PR 12-AUG-1999; 99US-0147303.  
 PR 13-AUG-1999; 99US-0147416.  
 PR 14-AUG-1999; 99US-0147439.  
 PR 15-AUG-1999; 99US-0147935.  
 PR 16-AUG-1999; 99US-0148171.  
 PR 17-AUG-1999; 99US-0148319.  
 PR 18-AUG-1999; 99US-0148341.  
 PR 19-AUG-1999; 99US-0148565.  
 PR 20-AUG-1999; 99US-0148684.  
 PR 21-AUG-1999; 99US-0149368.  
 PR 22-AUG-1999; 99US-0149375.  
 PR 23-AUG-1999; 99US-0149426.  
 PR 24-AUG-1999; 99US-0149722.  
 PR 25-AUG-1999; 99US-0149723.  
 PR 26-AUG-1999; 99US-0149929.  
 PR 27-AUG-1999; 99US-0149902.  
 PR 28-AUG-1999; 99US-0149930.  
 PR 29-AUG-1999; 99US-0150566.  
 PR 30-AUG-1999; 99US-0150884.  
 PR 31-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.  
 PR 28-AUG-1999; 99US-0151080.  
 PR 29-AUG-1999; 99US-0151303.  
 PR 30-AUG-1999; 99US-0151338.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 02-SEP-1999; 99US-0152363.  
 PR 03-SEP-1999; 99US-0153070.  
 PR 04-SEP-1999; 99US-0153758.  
 PR 05-SEP-1999; 99US-0154018.  
 PR 06-SEP-1999; 99US-0154039.  
 PR 07-SEP-1999; 99US-0154779.  
 PR 08-SEP-1999; 99US-0155139.  
 PR 09-SEP-1999; 99US-0155486.  
 PR 10-SEP-1999; 99US-0155659.  
 PR 11-SEP-1999; 99US-0156458.  
 PR 12-SEP-1999; 99US-0156596.  
 PR 13-SEP-1999; 99US-0157117.  
 PR 14-SEP-1999; 99US-0157753.  
 PR 15-SEP-1999; 99US-0157865.  
 PR 16-SEP-1999; 99US-0158029.  
 PR 17-SEP-1999; 99US-0158232.  
 PR 18-SEP-1999; 99US-0158369.  
 PR 19-SEP-1999; 99US-0159293.  
 PR 20-SEP-1999; 99US-0159294.  
 PR 21-SEP-1999; 99US-0159295.  
 PR 22-SEP-1999; 99US-0159329.  
 PR 23-SEP-1999; 99US-0159330.  
 PR 24-SEP-1999; 99US-0159331.  
 PR 25-SEP-1999; 99US-0159637.  
 PR 26-SEP-1999; 99US-0159638.  
 PR 27-SEP-1999; 99US-0159639.  
 PR 28-SEP-1999; 99US-0160741.  
 PR 29-SEP-1999; 99US-0160767.  
 PR 30-SEP-1999; 99US-0160768.  
 PR 01-OCT-1999; 99US-0160814.  
 PR 02-OCT-1999; 99US-0160815.  
 PR 03-OCT-1999; 99US-0160816.  
 PR 04-OCT-1999; 99US-0160980.  
 PR 05-OCT-1999; 99US-0160981.  
 PR 06-OCT-1999; 99US-0160982.  
 PR 07-OCT-1999; 99US-0160983.  
 PR 08-OCT-1999; 99US-0161404.  
 PR 09-OCT-1999; 99US-0161405.  
 PR 10-OCT-1999; 99US-0161406.  
 PR 11-OCT-1999; 99US-0161359.  
 PR 12-OCT-1999; 99US-0161360.  
 PR 13-OCT-1999; 99US-0161361.  
 PR 14-OCT-1999; 99US-0161362.  
 PR 15-OCT-1999; 99US-0161920.  
 PR 16-OCT-1999; 99US-0161921.  
 PR 17-OCT-1999; 99US-0161922.  
 PR 18-OCT-1999; 99US-0161923.  
 PR 19-OCT-1999; 99US-0162142.

Query Match 31.2% Score 5; DB 21; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
 DB 7 IDFI 11

RESULT 34

ABB09650  
 ID ABB09650 standard; Protein; 118 AA.

XX ABB09650;

XX 29-MAY-2002 (first entry)

DE Amino acid sequence of gene encoded mutacin I (MutA) peptide OrfZ.  
 DE Lantibiotic; mutacin I; mutA; Streptococcus mutans CH43; probiotic;  
 KW antimicrobial; multiple drug resistant bacteria; MutR; MutA; MutA;  
 KW MutB; MutC; MutD; MutP; MutT; OrfZ.

11-APR-2000; 2000FR-0004629.  
(INSP ) INST PASTEUR.  
Buchrieser C, Frangeul L, Couve E, Rusnlok C, Fsthi H, Dehoux P;  
Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend L;  
Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
Perez-Diaz J, Baquero F, Garcia del Portillo F, Gomez-Lopez N;  
Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
Rose M, Voss H;  
WPI; 2002-010914/01.  
Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
and prevention of *Listeria* and related bacterial infections, and  
related polypeptides -  
Claim 6; SEQ ID No 897; 192pp; French.  
The present invention relates to the genome sequence of *Listeria*  
*monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of  
it are useful for selecting probes and primers for detecting genes in *L.*  
*monocytogenes* and related organisms, and for studying genetic  
polymorphisms and other genomes. The present sequence is a protein  
expressed by the genome sequence of the present invention. Proteins  
expressed from the genome sequence are useful for raising specific  
antibodies, identification of *L. monocytogenes* and related organisms, and  
for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
B12. The genome sequence and proteins encoded by it are also useful for  
selecting compounds that regulate gene expression and cell replication  
and modulate *L. monocytogenes*-related diseases. In addition, the genome  
sequence and proteins encoded by it are useful in pharmaceutical and  
vaccine compositions for the treatment or prevention of infections by *L.*  
*monocytogenes* and related organisms.  
Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.  
Sequence 120 AA;  
Query Match 31.2%; Score 5; DB 23; Length 120;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 IDFII 11  
Db 27 IDFII 31  
RESULT 36  
ABP40387  
ID ABP40387 standard; Protein; 134 AA.  
XX AC ABP40387;  
XX DT 24-JUL-2002 (first entry)  
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5232.  
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX OS Antibacterial; gene therapy.  
XX OS Staphylococcus epidermidis.  
XX PN US6380370-B1.  
XX PD 30-APR-2002.  
XX PF 13-AUG-1998; 98US-0134001.  
XX SQ

Streptococcus mutans.  
US6342385-B1.  
29-JAN-2002.  
28-JUL-2000; 2000US-0627376.  
28-JUL-2000; 2000US-0627376.  
(UABR-) UAB RES FOUND.  
Qi F, Caulfield PW, Chen P;  
WPI; 2002-204553/26.  
New lantibiotic (mutacin) encoding gene, useful as a probiotic and as  
anti-microbial agents for treating multiple drug resistant bacteria  
such as certain strains of *Staphylococcus aureus*, which are known to be  
multiple drug resistant -  
Example; Columns 55-56; 37pp; English.  
ABB09645-50 represent gene encoded mutacin I (mutA) peptides. The  
mutacin I operon comprises eight genes in the order of MutR, MutA,  
MutA', MutB, MutC, MutD, MutP and MutT. MutR is the positive  
regulator for expression of the operon. MutA is believed to be the  
structural gene encoding premutacin I. MutB, MutC and MutD possibly  
constitute the modification apparatus for premutacin I, and MutP and  
MutT are the protease and ABC transporter, respectively, for  
transportation and processing of premutacin I. MutA is a lantibiotic,  
which is isolated from *Streptococcus mutans* CH43. The MutA protein has  
a molecular weight of approximately 2364 Da. MutA nucleic acid  
fragments are useful as probiotics and for producing the lantibiotic,  
mutacin I, in situ to protect subjects against pathogens or other  
undesirable organisms. The MutA nucleic acid and its encoded protein  
are particularly useful for as antimicrobial agents against, e.g.  
antibiotic resistant microorganisms. These are especially useful for  
treating multiple drug resistant bacteria such as certain strains of  
*Staphylococcus aureus*, which are known to be multiple drug resistant.  
Sequence 118 AA;  
Query Match 31.2%; Score 5; DB 23; Length 118;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 11 IFWIF 15  
Db 14 IFWIF 18  
RESULT 35  
ABB48192  
ID ABB48192 standard; Protein; 120 AA.  
XX AC ABB48192;  
XX DT 05-FEB-2002 (first entry)  
XX DE *Listeria monocytogenes* protein #896.  
XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
XX OS vitamin B12; bacterial infection; disease.  
XX OS *Listeria monocytogenes*.  
XX PN WO200177335-A2.  
XX PD 18-OCT-2001.  
XX PF 11-APR-2001; 2001WO-FR01118.

PR 14-AUG-1997; 97US-055779P.  
 XX 08-NOV-1997; 97US-064964P.  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX Doucette-Stamm LA, Bush D;  
 XX WPI: 2002-381255/41.  
 DR N-PSDB; ABN92932.  
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 PS Disclosure; SEQ ID 5232; 267pp; English.  
 XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX SQ Sequence 134 AA;  
 Query Match 31.2%; Score 5; DB 23; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 IDFTI 11  
 DB 89 IDFTI 93  
 RESULT 37  
 AAP60068  
 ID AAP60068 standard; Protein; 146 AA.  
 XX AC AAP60068;  
 XX DT 26-JUN-1991 (first entry)  
 XX DE Sequence of HTLV-III polypeptide 121 fusion protein.  
 XX KW AIDS; HIV; LAV; vaccine; diagnosis; antibody; fusion protein.  
 XX OS HTLV-III.  
 XX FH Key Location/Qualifiers  
 XX FT Region 1..41  
 XX FT Protein /note= "E.coli encoded"  
 XX FT Region 42..122  
 XX FT Region 123..146  
 XX FT /note= "vector encoded"  
 XX PN EPI199438-A.  
 XX PD 29-OCT-1986.  
 XX PF 03-MAR-1986; 86EP-0301488.  
 XX PR 01-MAR-1985; 85US-0707066.  
 XX PA (CHAN/) CHANG T W.  
 XX PA (CENT-) CENTOCOR INC.  
 XX PI Chang TW, Kato I, Chanda P, Chang NT, Putney SD;  
 XX

DR WPI; 1986-286179/44.  
 XX Human T-cell lymphotropic virus-III polypeptide 121 - useful in  
 PT assays to detect AIDS and for vaccination  
 XX Claim 3; Page 14; 17pp; English.  
 XX The HTLV-III DNA segment of clone 121 was excised from pMR100 and  
 CC sequenced (AAN60081). Based upon the DNA sequence, the putative AA  
 CC order of the HTLV-III polypeptide could be assigned (AAP60067). In  
 CC order to enhance expression the HTLV-III polypeptide in E.coli the  
 CC HTLV-III segment of clone 121 was cloned into a high expression  
 CC vector; E.coli transformed with the recombinant vector expressed a  
 CC 15 kD fusion protein with short fusion partners at both ends (41 AA  
 CC of E.coli, the 83 AA residues encoding by HTLV-III polypeptide and  
 CC 13 AA residues encoding by AAP60068).  
 XX SQ Sequence 146 AA;  
 Query Match 31.2%; Score 5; DB 7; Length 146;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 NCGID 8  
 DB 26 NCGID 30  
 RESULT 38  
 AAB92462  
 ID AAB92462 standard; Protein; 148 AA.  
 XX AC AAB92462;  
 XX DT 26-JUN-2001 (first entry)  
 XX DE Human protein sequence SEQ ID NO:10515.  
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX OS Homo sapiens.  
 XX PN EPI074617-A2.  
 XX PD 07-FEB-2001.  
 XX PF 28-JUL-2000; 2000EP-0116126.  
 XX PR 29-JUL-1999; 99JP-0248036.  
 XX PR 27-AUG-1999; 99JP-0300253.  
 XX PR 11-JAN-2000; 2000JP-0118776.  
 XX PR 02-MAY-2000; 2000JP-0183767.  
 XX PR 09-JUN-2000; 2000JP-0241899.  
 XX PA (HELI-) HELIX RES INST.  
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX Claim 8; SEQ ID 10515; 2537pp + CD ROM; English.  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of



SQ Sequence 150 AA;

Query Match 31.2%; Score 5; DB 18; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 Db 111 GIDFI 115

RESULT 41

ABG30352  
 ID ABG30352 standard; Protein; 151 AA.

AC ABG30352;

DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #30343.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

N-PSDB; AAS94539.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

PS Claim 20; SEQ ID No 60711; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 151 AA;

Query Match

Best Local Similarity 31.2%; Score 5; DB 22; Length 151;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9  
 Db 144 CGIDF 148

RESULT 42

ABG22986  
 ID ABG22986 standard; Protein; 152 AA.

AC ABG22986;

DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #22977.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

N-PSDB; AAS87173.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

PS Claim 20; SEQ ID No 53345; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 152 AA;

Query Match

Best Local Similarity 31.2%; Score 5; DB 22; Length 152;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Tue, May 13 12:12:54 2003

us-09-854-133-587.olog.rag

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 FWIEW 16  
 ID IIIII  
 Db 115 FWIEW 119

## RESULT 43

ABP29925 standard; Protein; 157 AA.

XX AC ABP29925;  
 XX AC

02-JUL-2002 (first entry)

Streptococcus polypeptide SEQ ID NO 9026.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 group A streptococcus; Streptococcus pyogenes; antibacterial;  
 antiinflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus agalactiae.

WO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB04789.

27-OCT-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0028727.

07-MAR-2001; 2001GB-0005640.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;  
 Tettelin H;

WPI; 2002-352536/38.

N-PSDB; ABN70556.

New Streptococcus protein for the treatment or prevention of infection  
 or disease caused by Streptococcus bacteria, such as meningitis, and  
 for detecting a compound that binds to the protein -

Claim 1; Page 4024; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B  
 streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 the specification. The proteins have antibacterial and antiinflammatory  
 activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 antibodies that bind (I) are used in the manufacture of medicaments for  
 the treatment or prevention of infection or disease caused by  
 Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 Nucleic acids encoding (I) are used to detect Streptococcus in a  
 biological sample. (I) is used to determine whether a compound binds to  
 used as a vaccine or diagnostic composition. The disease caused by  
 Streptococcus that is prevented or treated may be meningitis. Nucleic  
 acid encoding (I) may be used to recombinantly produce (I) and may be  
 used in gene therapy. Antibodies to (I) are used for affinity  
 chromatography, immunoassays, and distinguishing/identifying  
 Streptococcus proteins.

Sequence 157 AA;

Query Match 31.2%; Score 5; DB 23; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11

Db 34 IDFII 38  
 IIIII

## RESULT 44

ABP30763 standard; Protein; 157 AA.

XX AC ABP30763;  
 XX AC

02-JUL-2002 (first entry)

Streptococcus polypeptide SEQ ID NO 10702.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 group A streptococcus; Streptococcus pyogenes; antibacterial;  
 antiinflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus agalactiae.

WO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB04789.

27-OCT-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0028727.

07-MAR-2001; 2001GB-0005640.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;  
 Tettelin H;

WPI; 2002-352536/38.

N-PSDB; ABN71394.

New Streptococcus protein for the treatment or prevention of infection  
 or disease caused by Streptococcus bacteria, such as meningitis, and  
 for detecting a compound that binds to the protein -

Claim 1; Page 4179; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B  
 streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 the specification. The proteins have antibacterial and antiinflammatory  
 activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 antibodies that bind (I) are used in the manufacture of medicaments for  
 the treatment or prevention of infection or disease caused by  
 Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 Nucleic acids encoding (I) are used to detect Streptococcus in a  
 biological sample. (I) is used to determine whether a compound binds to  
 used as a vaccine or diagnostic composition. The disease caused by  
 Streptococcus that is prevented or treated may be meningitis. Nucleic  
 acid encoding (I) may be used to recombinantly produce (I) and may be  
 used in gene therapy. Antibodies to (I) are used for affinity  
 chromatography, immunoassays, and distinguishing/identifying  
 Streptococcus proteins.

Sequence 157 AA;

Query Match 31.2%; Score 5; DB 23; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11

Db 34 IDFII 38  
 IIIII

Job time : 11.3451 secs

RESULT 45  
 ABG12807  
 ID ABG12807 standard; Protein; 160 AA.  
 XX AC ABG12807;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #12798.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX DR WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS76994.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 Claim 20; SEQ ID NO 43166; 103pp; English.  
 The invention relates to isolated polynucleotide (I) and  
 polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 and gene mapping, and in recombinant production of (II). The  
 polynucleotides are also used in diagnostics as expressed sequence tags  
 for identifying expressed genes. (I) is useful in gene therapy techniques  
 to restore normal activity of (II) or to treat disease states involving  
 (II). (II) is useful for generating antibodies against it, detecting or  
 quantitating a polypeptide in tissue, as molecular weight markers and as  
 a food supplement. (II) and its binding partners are useful in medical  
 imaging of sites expressing (II). (I) and (II) are useful for treating  
 disorders involving aberrant protein expression or biological activity.  
 The polypeptide and polynucleotide sequences have applications in  
 diagnostics, forensics, gene mapping, identification of mutations  
 responsible for genetic disorders or other traits to assess biodiversity  
 and to produce other types of data and products dependent on DNA and  
 amino acid sequences. ABG00010-ABG30377 represent novel human  
 diagnostic amino acid sequences of the invention.  
 Note: The sequence data for this patent did not appear in the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.  
 SQ Sequence 160 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 IIFWI 14  
 Db 139 IIFWI 143

Search completed: May 11, 2003, 20:10:50



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:09:32 ; Search time 3.9646 seconds  
(without alignments)  
118.743 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANGIDFIIFWIFW 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	6	37.5	399	1 US-08-414-926A-5	Sequence 5, Appli
2	6	37.5	399	2 US-08-926-922-5	Sequence 5, Appli
3	6	37.5	399	3 US-09-253-682-5	Sequence 5, Appli
4	6	37.5	399	4 US-09-527-657-5	Sequence 5, Appli
5	5	31.2	69	1 US-08-410-804-5	Sequence 5, Appli
6	5	31.2	69	1 US-08-259-514-5	Sequence 5, Appli
7	5	31.2	69	2 US-08-858-311-5	Sequence 5, Appli
8	5	31.2	74	4 US-09-134-001C-5666	Sequence 5666, Ap
9	5	31.2	103	2 US-08-529-878B-41	Sequence 41, Appl
10	5	31.2	118	4 US-09-627-376-17	Sequence 17, Appl
11	5	31.2	134	4 US-09-134-001C-5232	Sequence 5232, Ap
12	5	31.2	220	3 US-08-228-208A-21	Sequence 21, Appl
13	5	31.2	223	1 US-08-505-058-5	Sequence 5, Appli
14	5	31.2	223	2 US-08-459-818-25	Sequence 25, Appl
15	5	31.2	223	2 US-08-889-666-25	Sequence 25, Appl
16	5	31.2	223	2 US-08-465-078-25	Sequence 25, Appl
17	5	31.2	223	2 US-08-725-776-25	Sequence 25, Appl
18	5	31.2	223	2 US-08-488-062-25	Sequence 25, Appl
19	5	31.2	229	4 US-09-134-001C-2965	Sequence 2965, Ap
20	5	31.2	238	4 US-08-858-207A-420	Sequence 420, App
21	5	31.2	270	3 US-09-188-579-84	Sequence 84, Appl
22	5	31.2	270	4 US-09-315-444-84	Sequence 84, Appl
23	5	31.2	270	4 US-09-721-362-84	Sequence 84, Appl
24	5	31.2	295	1 US-08-118-270-79	Sequence 79, Appl
25	5	31.2	295	5 PCR-US93-0852B-79	Sequence 79, Appl
26	5	31.2	401	3 US-08-974-022-2	Sequence 2, Appli
27	5	31.2	401	4 US-08-795-445A-2	Sequence 2, Appli

28	5	31.2	401	4 US-08-795-447A-2	Sequence 2, Appli
29	5	31.2	401	4 US-08-974-186-2	Sequence 2, Appli
30	5	31.2	401	4 US-08-795-446B-2	Sequence 2, Appli
31	5	31.2	401	4 US-08-706-945D-124	Sequence 124, App
32	5	31.2	518	4 US-09-134-001C-4451	Sequence 4451, Ap
33	5	31.2	580	4 US-08-913-159-10	Sequence 10, Appl
34	5	31.2	589	1 US-08-399-646-2	Sequence 2, Appli
35	5	31.2	589	1 US-08-607-321-2	Sequence 2, Appli
36	5	31.2	589	2 US-08-961-240-2	Sequence 2, Appli
37	5	31.2	589	2 US-08-605-501-2	Sequence 2, Appli
38	5	31.2	596	1 US-08-399-646-12	Sequence 12, Appl
39	5	31.2	596	1 US-08-607-321-12	Sequence 12, Appl
40	5	31.2	596	2 US-08-961-240-12	Sequence 12, Appl
41	5	31.2	596	2 US-08-605-501-12	Sequence 12, Appl
42	5	31.2	600	6 5268463-2	Patent No. 5268463
43	5	31.2	602	2 US-08-882-704A-5	Sequence 5, Appli
44	5	31.2	602	4 US-09-151-957-5	Sequence 5, Appli
45	5	31.2	602	6 5432081-2	Patent No. 5432081
46	5	31.2	603	4 US-09-149-727-6	Sequence 6, Appli
47	5	31.2	832	3 US-08-630-820-7	Sequence 7, Appli
48	5	31.2	1242	4 US-09-488-270A-2	Sequence 2, Appli
49	5	31.2	2763	3 US-08-477-451-15	Sequence 15, Appl
50	5	31.2	2763	3 US-08-496-944-2	Sequence 2, Appli
51	4	25.0	8	4 US-09-187-859-1001	Sequence 1001, Ap
52	4	25.0	8	4 US-09-187-859-1322	Sequence 1322, Ap
53	4	25.0	8	4 US-09-187-859-1903	Sequence 1903, Ap
54	4	25.0	8	4 US-09-187-859-3225	Sequence 3225, Ap
55	4	25.0	9	4 US-09-187-859-2885	Sequence 2885, Ap
56	4	25.0	9	4 US-09-187-859-3226	Sequence 3226, Ap
57	4	25.0	9	4 US-09-187-859-3897	Sequence 3897, Ap
58	4	25.0	9	4 US-09-187-859-3963	Sequence 3963, Ap
59	4	25.0	10	4 US-09-187-859-3010	Sequence 3010, Ap
60	4	25.0	10	4 US-09-187-859-3227	Sequence 3227, Ap
61	4	25.0	10	4 US-09-187-859-3902	Sequence 3902, Ap
62	4	25.0	10	4 US-09-187-859-3968	Sequence 3968, Ap
63	4	25.0	13	4 US-09-258-754-108	Sequence 108, App
64	4	25.0	13	4 US-09-042-107-108	Sequence 108, App
65	4	25.0	15	4 US-09-009-953-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1  
US-08-414-926A-5  
; Sequence 5, Application US/08414926A  
; Patent No. 5721354  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
; STREET: 5 Palo Alto Square  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-2155  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/414,926A  
; FILING DATE: March 31, 1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS  
; TELECOMMUNICATION INFORMATION:

```
;
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-414-926A-5

Query Match 37.5%; Score 6; DB 1; Length 399;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14
Db 20 FIIFWI 25

RESULT 2
US-08-926-922-5
; Sequence 5, Application US/08926922
; Patent No. 5925751
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,922
; FILING DATE: September 10, 1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-926-922-5

Query Match 37.5%; Score 6; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14
Db 20 FIIFWI 25

RESULT 3
US-09-253-682-5
; Sequence 5, Application US/09253682
; Patent No. 6040170
; GENERAL INFORMATION:
```

```
;
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/253,682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/926,922
; FILING DATE: September 10, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-253-682-5

Query Match 37.5%; Score 6; DB 3; Length 399;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14
Db 20 FIIFWI 25

RESULT 4
US-09-527-657-5
; Sequence 5, Application US/09527657
; Patent No. 6291236
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/527,657
; FILING DATE: 17-Mar-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: US/08/926,922  
FILING DATE: September 10, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: AVIR 11A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-834-1448  
TELEFAX: 510-839-7810  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 399 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-527-657-5

Query Match 37.5%; Score 6; DB 4; Length 399;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14  
Db 20 FIIFWI 25

RESULT 5  
US-08-410-804-5  
Sequence 5, Application US/08410804  
Patent No. 5632994  
GENERAL INFORMATION:  
APPLICANT: Reed, John C.  
APPLICANT: Sato, Takaaki  
TITLE OF INVENTION: FAS ASSOCIATED PROTEINS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cathryn Campbell  
STREET: 4370 La Jolla Village Drive, Ste 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/410,804  
FILING DATE: 27-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/259,514  
FILING DATE: 14-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 1389  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 69 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-410-804-5

Query Match 31.2%; Score 5; DB 1; Length 69;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
Db 58 IDFII 62

RESULT 6  
US-08-259-514-5  
Sequence 5, Application US/08259514  
Patent No. 5747245  
GENERAL INFORMATION:  
APPLICANT: Reed, John C.  
APPLICANT: Sato, Takaaki  
TITLE OF INVENTION: FAS ASSOCIATED PROTEINS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cathryn Campbell  
STREET: 4370 La Jolla Village Drive, Ste 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/259,514  
FILING DATE: 14-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 9954  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 69 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-259-514-5

Query Match 31.2%; Score 5; DB 1; Length 69;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
Db 58 IDFII 62

RESULT 7  
US-08-858-311-5  
Sequence 5, Application US/08858311  
Patent No. 5876939  
GENERAL INFORMATION:  
APPLICANT: Reed, John C.  
APPLICANT: Sato, Takaaki  
TITLE OF INVENTION: FAS ASSOCIATED PROTEINS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cathryn Campbell  
STREET: 4370 La Jolla Village Drive, Ste 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,311
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/410,804
; FILING DATE: 27-MAR-1995
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1389
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 69 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-858-311-5

```

```

Query Match          31.2%; Score 5; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 IDFII 11
    |||||
DB 58 IDFII 62

```

```

RESULT 8
US-09-134-001C-5666
; Sequence 5666, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5666
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5666

```

```

Query Match          31.2%; Score 5; DB 4; Length 74;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 9 FIIFW 13
    |||||
DB 40 FIIFW 44

```

```

RESULT 9
US-08-529-878B-41
; Sequence 41, Application US/08529878B
; Patent No. 5932556
; GENERAL INFORMATION:
; APPLICANT: Tam, Robert C.

```

```

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: REGULATION OF CD28 EXPRESSION
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Crockett & Fish
; STREET: 3000 S. Augusta Court
; CITY: La Habra
; STATE: California
; COUNTRY: United States of America
; ZIP: 90631
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/529,878B
; FILING DATE: 13-SEP-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Fish, Robert D.
; REGISTRATION NUMBER: 33,880
; REFERENCE/DOCKET NUMBER: 213/003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-525-3433
; TELEFAX: 714-525-3303
; TELEX:
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-529-878B-41

```

```

Query Match          31.2%; Score 5; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 9 FIIFW 13
    |||||
DB 57 FIIFW 61

```

```

RESULT 10
US-09-627-376-17
; Sequence 17, Application US/09627376
; Patent No. 6342385
; GENERAL INFORMATION:
; APPLICANT: Qi, Fengxia
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS
; FILE REFERENCE: UAB-17402/22
; CURRENT APPLICATION NUMBER: US/09/627,376
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-627-376-17

```

```

Query Match          31.2%; Score 5; DB 4; Length 118;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 11 IFWIF 15
    |||||
DB 14 IFWIF 18

```

```

RESULT 11
US-09-134-001C-5232

```

; Sequence 5232, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5232  
; LENGTH: 134  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5232

Query Match 31.2%; Score 5; DB 4; Length 134;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 IDFI 11  
Db 89 IDFI 93

RESULT 12  
US-08-228-208A-21  
; Sequence 21, Application US/08228208A  
; Patent No. 6090914  
; GENERAL INFORMATION:  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Damle, Nitin K.  
; APPLICANT: Brady, William  
; APPLICANT: Wallace, Philip M.  
; TITLE OF INVENTION: CTIA4/CD28Ig HYBRID FUSION  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Boulevard, Suite 400  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/228,208A  
; FILING DATE: 15-APR-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/008,898  
; FILING DATE: 22-JAN-1993  
; APPLICATION NUMBER: 07/723,617  
; FILING DATE: 27-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436-30US01  
; TELEPHONE: 310 445-1140  
; TELEFAX: 310 445-9031  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 220 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-228-208A-21

Query Match 31.2%; Score 5; DB 3; Length 220;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FIIFW 13  
Db 174 FIIFW 178

RESULT 13  
US-08-505-058-5  
; Sequence 5, Application US/08505058  
; Patent No. 5773253  
; GENERAL INFORMATION:  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Peach, Robert  
; TITLE OF INVENTION: CTIA4 Mutant Molecules and Uses Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/505,058  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/228,208  
; FILING DATE: 15-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436.30US11  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-505-058-5

Query Match 31.2%; Score 5; DB 1; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FIIFW 13  
Db 176 FIIFW 180

RESULT 14  
US-08-459-818-25  
; Sequence 25, Application US/08459818

; Patent No. 5851795  
; GENERAL INFORMATION:  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Damle, Nitin K.  
; APPLICANT: Brady, William  
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: FastSeq 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/459,818  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436.35US02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-459-818-25  
Query Match 31.2%; Score 5; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 FIIFW 13  
Db 176 FIIFW 180  
RESULT 15  
US-08-889-666-25  
; Sequence 25, Application US/08889666  
; Patent No. 5885579  
; GENERAL INFORMATION:  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Damle, Nitin K.  
; APPLICANT: Brady, William  
; APPLICANT: Kiener, Peter A.  
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/889,666  
; FILING DATE: 08-JUL-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/375390  
; FILING DATE: 18-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436-35US01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-889-666-25  
Query Match 31.2%; Score 5; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 FIIFW 13  
Db 176 FIIFW 180  
RESULT 16  
US-08-465-078-25  
; Sequence 25, Application US/08465078  
; Patent No. 5885796  
; GENERAL INFORMATION:  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Damle, Nitin K.  
; APPLICANT: Brady, William  
; APPLICANT: Kiener, Peter A.  
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,078  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/375390  
; FILING DATE: 18-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436-35US01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-465-078-25

Query Match 31.2%; Score 5; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIFW 13  
|  
|  
|  
|  
Db 176 FIFW 180

## RESULT 17

US-08-725-776-25  
; Sequence 25, Application US/08725776  
; Patent No. 5968510  
; GENERAL INFORMATION:

; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Damle, Nitin K.  
; APPLICANT: Brady, William  
; APPLICANT: Kiener, Peter A.  
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90025

; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/725,776  
; FILING DATE:  
; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/375390  
; FILING DATE: 18-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436-35US01  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-725-776-25

Query Match 31.2%; Score 5; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIFW 13  
|  
|  
|  
|  
Db 176 FIFW 180

## RESULT 18

US-08-488-062-25  
; Sequence 25, Application US/08488062  
; Patent No. 5977318  
; GENERAL INFORMATION:  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Damle, Nitin K.  
; APPLICANT: Brady, William  
; APPLICANT: Kiener, Peter A.  
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90025

; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,062  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/375390  
; FILING DATE: 18-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436-35US01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-488-062-25

Query Match 31.2%; Score 5; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIFW 13  
|  
|  
|  
|  
Db 176 FIFW 180

## RESULT 19

US-09-134-001C-2965  
; Sequence 2965, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2965  
; LENGTH: 229

```
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2965

Query Match          31.2%; Score 5; DB 4; Length 229;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14
Db 63 IIFWI 67

RESULT 20
US-08-858-207A-420
; Sequence 420, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 420:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-420

Query Match          31.2%; Score 5; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 32 GIDFI 36

RESULT 21
US-09-188-579-84
; Sequence 84, Application US/09188579B
```

```
; Patent No. 6107040
; GENERAL INFORMATION:
; APPLICANT: Shuman, Stewart
; TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
; FILE REFERENCE: D6185
; CURRENT APPLICATION NUMBER: US/09/188,579B
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 114
; SEQ ID NO 84
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Chlorella virus PBCV-1
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of RNA guanylyltransferase.
US-09-188-579-84

Query Match          31.2%; Score 5; DB 3; Length 270;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11
Db 188 IDFII 192

RESULT 22
US-09-315-444-84
; Sequence 84, Application US/09315444A
; Patent No. 6232070
; GENERAL INFORMATION:
; APPLICANT: Shuman, Stewart
; TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
; FILE REFERENCE: D6185CIP
; CURRENT APPLICATION NUMBER: US/09/315,444A
; CURRENT FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: US 09/188,579
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 84
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Chlorella virus PBCV-1
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of RNA guanylyltransferase.
US-09-315-444-84

Query Match          31.2%; Score 5; DB 4; Length 270;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11
Db 188 IDFII 192

RESULT 23
US-09-721-362-84
; Sequence 84, Application US/09721362
; Patent No. 6420163
; GENERAL INFORMATION:
; APPLICANT: Shuman, Stewart
; TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
; FILE REFERENCE: D6185CIP/D
; CURRENT APPLICATION NUMBER: US/09/721,362
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 09/315,444
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 84
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Chlorella virus PBCV-1
; FEATURE:
```



; OTHER INFORMATION: Amino acid sequence of RNA guanylyltransferase.  
US-09-721-362-84

Query Match 31.2%; Score 5; DB 4; Length 270;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 IDFI 11  
Db 188 IDFI 192

## RESULT 24

US-08-118-270-79  
; Sequence 79, Application US/08118270  
; Patent No. 5508384  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Randall B.  
; APPLICANT: Schuster, David I.  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/118,270  
; FILING DATE: 09-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/943,236  
; FILING DATE: 10-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Townsend, Kevin G.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: MURPHY-2A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633

; INFORMATION FOR SEQ ID NO: 79:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 295 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-118-270-79

Query Match 31.2%; Score 5; DB 1; Length 295;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FIIFW 13  
Db 207 FIIFW 211

## RESULT 25

PCT-US93-08528-79  
; Sequence 79, Application PC/TUS9308528  
; GENERAL INFORMATION:  
; APPLICANT: New York University  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF

; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/08528  
; FILING DATE: 09-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/943,236  
; FILING DATE: 10-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Townsend, Kevin G.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: MURPHY-2 PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 79:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 295 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
PCT-US93-08528-79

Query Match 31.2%; Score 5; DB 5; Length 295;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FIIFW 13  
Db 207 FIIFW 211

## RESULT 26

US-08-974-022-2  
; Sequence 2, Application US/08974022  
; Patent No. 6015938  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Behavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/974,022  
; FILING DATE: 12-DEC-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/577,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 401 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-974-022-2

Query Match 31.2%; Score 5; DB 3; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
Db 194 NCGID 198

RESULT 27  
US-08-795-445A-2  
Sequence 2, Application US/08795445A  
Patent No. 6284485  
GENERAL INFORMATION:  
APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,445A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/577,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 401 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-795-445A-2

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
Db 194 NCGID 198

RESULT 28

US-08-795-447A-2  
Sequence 2, Application US/08795447A  
Patent No. 6284728  
GENERAL INFORMATION:  
APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: Osteoprotegerin  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: One Amgen Center Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91362-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,447A  
FILING DATE:

CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378D2  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 401 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-795-447A-2

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
Db 194 NCGID 198

RESULT 29  
US-08-974-186-2  
Sequence 2, Application US/08974186  
Patent No. 6284740  
GENERAL INFORMATION:  
APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,186  
FILING DATE:  
CLASSIFICATION:

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/577,788  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Winter, Robert B.  
;; REFERENCE/DOCKET NUMBER: A-378  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 401 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-974-186-2

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
|||||  
DB 194 NCGID 198

RESULT 30  
US-08-795-446B-2  
; Sequence 2, Application US/08795446B  
; Patent No. 6288032

;; GENERAL INFORMATION:  
;; APPLICANT: Boyle, William J.  
;; APPLICANT: Lacey, David L.  
;; APPLICANT: Calzone, Frank J.  
;; APPLICANT: Chang, Ming-Shi  
;; TITLE OF INVENTION: OSTEOPROTEGERIN  
;; NUMBER OF SEQUENCES: 53  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Amgen Inc.  
;; STREET: 1840 Dehavilland Drive  
;; CITY: Thousand Oaks  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 91320-1789

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/795,446B  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/577,788  
;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Winter, Robert B.  
;; REFERENCE/DOCKET NUMBER: A-378  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 401 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-795-446B-2

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
|||||  
DB 194 NCGID 198

RESULT 31  
US-08-706-945D-124  
; Sequence 124, Application US/08706945D  
; Patent No. 6369027

;; GENERAL INFORMATION:  
;; APPLICANT: Boyle, William  
;; APPLICANT: Lacey, David  
;; APPLICANT: Calzone, Frank  
;; APPLICANT: Chang, Ming-Shi  
;; TITLE OF INVENTION: Osteoprotegerin  
;; FILE REFERENCE: A-378CIP  
;; CURRENT APPLICATION NUMBER: US/08/706,945D  
;; CURRENT FILING DATE: 1996-09-03  
;; PRIOR APPLICATION NUMBER: 08/577,788  
;; PRIOR FILING DATE: 1995-12-22  
;; NUMBER OF SEQ ID NOS: 145  
;; SOFTWARE: Patent version 3.1  
;; SEQ ID NO 124  
;; LENGTH: 401  
;; TYPE: PRT  
;; ORGANISM: Rattus rattus  
US-08-706-945D-124

Query Match 31.2%; Score 5; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
|||||  
DB 194 NCGID 198

RESULT 32  
US-09-134-001C-4451  
; Sequence 4451, Application US/09134001C  
; Patent No. 6380370

;; GENERAL INFORMATION:  
;; APPLICANT: Lynn Doucette-Stamm et al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
;; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: GFC-007  
;; CURRENT APPLICATION NUMBER: US/09/134,001C  
;; CURRENT FILING DATE: 1998-08-13  
;; PRIOR APPLICATION NUMBER: US 60/064,964  
;; PRIOR FILING DATE: 1997-11-08  
;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 4451  
;; LENGTH: 518  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4451

Query Match 31.2%; Score 5; DB 4; Length 518;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
|||||  
DB 141 GIDFI 145

RESULT 33  
US-08-913-159-10  
; Sequence 10, Application US/08913159  
; Patent No. 6300109

;; GENERAL INFORMATION:  
;; APPLICANT:  
;; TITLE OF INVENTION: Plasmid-derived type II  
;; TITLE OF INVENTION: restriction-modification systems from Lactococcus lactis  
;; NUMBER OF SEQUENCES: 14  
;; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,159
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0179/95
; FILING DATE: 17-FEB-1995
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-913-159-10

Query Match          31.2%; Score 5; DB 4; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11
Db 148 IDFI 152

RESULT 34
US-08-399-646-2
; Sequence 2, Application US/08399646
; Patent No. 5536781
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: HATTORI, Kazuko
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,646
; FILING DATE: 07-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59834
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59840
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-399-646-2

Query Match          31.2%; Score 5; DB 1; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 138 GIDFI 142

RESULT 35
US-08-607-321-2
; Sequence 2, Application US/08607321
; Patent No. 5716813
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: HATTORI, Kazuko
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/607,321
; FILING DATE: 26-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,646
; FILING DATE: 07-MAR-1995
; APPLICATION NUMBER: JP 59834
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59840
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-607-321-2

Query Match          31.2%; Score 5; DB 1; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 138 GIDFI 142
```

RESULT 36  
US-08-961-240-2  
; Sequence 2, Application US/08961240  
; Patent No. 5830715  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,240  
; FILING DATE: 30-OCT-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 59840  
; FILING DATE: 07-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA-5  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-961-240-2  
Query Match 31.2%; Score 5; DB 2; Length 589;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFI 10  
Db 138 GIDFI 142  
RESULT 37  
US-08-605-501-2  
; Sequence 2, Application US/08605501  
; Patent No. 5834287  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,501  
; FILING DATE: 26-FEB-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 59840  
; FILING DATE: 07-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA-5  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-399-646-12  
Query Match 31.2%; Score 5; DB 2; Length 589;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFI 10  
Db 138 GIDFI 142  
RESULT 38  
US-08-399-646-12  
; Sequence 12, Application US/08399646  
; Patent No. 5556781  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,501  
; FILING DATE: 26-FEB-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 59840  
; FILING DATE: 07-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA-5  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-605-501-2  
Query Match 31.2%; Score 5; DB 2; Length 589;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFI 10  
Db 138 GIDFI 142

;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/399,646  
;; FILING DATE: 07-MAR-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA: JP 59834  
;; FILING DATE: 07-MAR-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 59840  
;; FILING DATE: 07-MAR-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BROWDY, Roger L.  
;; REGISTRATION NUMBER: 25,618  
;; REFERENCE/DOCKET NUMBER: KUBOTA-5  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 596 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-399-646-12

Query Match 31.2%; Score 5; DB 1; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.5e-02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 145 GIDFI 149

;; RESULT 39  
;; US-08-607-321-12  
;; Sequence 12, Application US/08607321  
;; Patent No. 5716813  
;; GENERAL INFORMATION:  
;; APPLICANT: KUBOTA, Michio  
;; APPLICANT: TSUSAKI, Keiji  
;; APPLICANT: HATTORI, Kazuko  
;; APPLICANT: SUGIMOTO, Toshiyuki  
;; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
;; ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BROWDY AND NEIMARK  
;; STREET: 419 Seventh Street, N.W., Suite 300  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20004  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/607,321  
;; FILING DATE: 26-FEB-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/399,646  
;; FILING DATE: 07-MAR-1995  
;; APPLICATION NUMBER: JP 59834  
;; FILING DATE: 07-MAR-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 59840  
;; FILING DATE: 07-MAR-1994  
;; ATTORNEY/AGENT INFORMATION:

;; NAME: BROWDY, Roger L.  
;; REGISTRATION NUMBER: 25,618  
;; REFERENCE/DOCKET NUMBER: KUBOTA-5  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 596 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-607-321-12

Query Match 31.2%; Score 5; DB 1; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 145 GIDFI 149

;; RESULT 40  
;; US-08-961-240-12  
;; Sequence 12, Application US/08961240  
;; Patent No. 5830715  
;; GENERAL INFORMATION:  
;; APPLICANT: KUBOTA, Michio  
;; APPLICANT: TSUSAKI, Keiji  
;; APPLICANT: HATTORI, Kazuko  
;; APPLICANT: SUGIMOTO, Toshiyuki  
;; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
;; ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BROWDY AND NEIMARK  
;; STREET: 419 Seventh Street, N.W., Suite 300  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20004  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/961,240  
;; FILING DATE: 30-OCT-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/399,646  
;; FILING DATE: 07-MAR-1995  
;; APPLICATION NUMBER: JP 59834  
;; FILING DATE: 07-MAR-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 59840  
;; FILING DATE: 07-MAR-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BROWDY, Roger L.  
;; REGISTRATION NUMBER: 25,618  
;; REFERENCE/DOCKET NUMBER: KUBOTA-5  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 596 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear

; MOLECULE TYPE: protein  
US-08-961-240-12

Query Match 31.2%; Score 5; DB 2; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 GIDFI 10  
|||||  
Db 145 GIDFI 149

## RESULT 41

US-08-605-501-12  
; Sequence 12, Application US/08605501  
; Patent No. 5834287

## ; GENERAL INFORMATION:

; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,501  
; FILING DATE: 26-FEB-1996

## ; CLASSIFICATION: 435

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 59840

## ; FILING DATE: 07-MAR-1994

## ; ATTORNEY/AGENT INFORMATION:

; NAME: BROWDY, Roger L.

## ; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: KUBOTA-5

## ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; TELEX: 248633

## ; INFORMATION FOR SEQ ID NO: 12:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 596 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-605-501-12

Query Match 31.2%; Score 5; DB 2; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 GIDFI 10  
|||||  
Db 145 GIDFI 149

## RESULT 42

5268463-2

; Patent No. 5268463

; APPLICANT: JEFFERSON, RICHARD A.

; TITLE OF INVENTION: PLANT PROMOTER a-GLUCURONIDASE GENE

## ; CONSTRUCT

; NUMBER OF SEQUENCES: 9

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/447,976

; FILING DATE: 08-DEC-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 119,102

; FILING DATE: 10-NOV-1987

; APPLICATION NUMBER: 264,586

; FILING DATE: 31-OCT-1988

; SEQ ID NO: 2:

; LENGTH: 600

5268463-2

Query Match 31.2%; Score 5; DB 6; Length 600;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 NCGID 8

|||||

Db 27 NCGID 31

## RESULT 43

US-08-882-704A-5

; Sequence 5, Application US/08882704A

; Patent No. 5879906

## ; GENERAL INFORMATION:

; APPLICANT: Jefferson, Richard A.

; APPLICANT: Wilson, Katherine J.

; APPLICANT: Leader, Michael

; TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

## ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/882,704A

; FILING DATE: 25-JUN-1997

; CLASSIFICATION: 435

## ; ATTORNEY/AGENT INFORMATION:

; NAME: No. 5879906tenburg Ph.D., Carol

; REGISTRATION NUMBER: 39,317

; REFERENCE/DOCKET NUMBER: 190106.404

## ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 602 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-882-704A-5

Query Match 31.2%; Score 5; DB 2; Length 602;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
|||||  
Db 27 NCGID 31

## RESULT 44

US-09-151-957-5  
; Sequence 5, Application US/09151957  
; Patent No. 6429292  
; GENERAL INFORMATION:  
; APPLICANT: Jefferson, Richard A.  
; ; Wilson, Katherine J.  
; ; Leader, Michael  
; ;  
; TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/151,957  
; FILING DATE: 11-Sep-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/882,704  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6429292tenburg Ph.D., Carol  
; REGISTRATION NUMBER: 39,317  
; REFERENCE/DOCKET NUMBER: 190106.404  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 602 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-151-957-5

Query Match 31.2%; Score 5; DB 4; Length 602;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
|||||  
Db 27 NCGID 31

## RESULT 45

5432081-2  
; Patent No. 5432081  
; APPLICANT: JEFFERSON, RICHARD A.  
; TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI  
; GLUCURONIDE PERMEASE GENE  
; NUMBER OF SEQUENCES: 10  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/138,546  
; FILING DATE: 15-OCT-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 447,976  
; FILING DATE: 08-DEC-1989  
; APPLICATION NUMBER: 264,586



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:12:57 ; Search time 5.23894 Seconds  
(without alignments)  
281.051 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FOANCGIDFIIFWIFW 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	16	9	US-09-854-133-587
2	16	100.0	16	10	US-09-738-973-587
3	16	100.0	97	9	US-09-854-133-586
4	16	100.0	97	10	US-09-738-973-586
5	5	31.2	41	9	US-09-820-843A-43
6	5	31.2	44	10	US-09-864-761-39223
7	5	31.2	67	9	US-10-073-961-221
8	5	31.2	67	10	US-09-764-887-221
9	5	31.2	68	9	US-10-102-806-693
10	5	31.2	82	9	US-09-764-891-3247
11	5	31.2	104	9	US-09-764-891-3077
12	5	31.2	118	12	US-10-047-676A-17
13	5	31.2	166	9	US-09-738-626-4531
14	5	31.2	166	9	US-09-746-660A-94
15	5	31.2	166	9	US-09-746-660A-96
16	5	31.2	198	9	US-09-791-279-200
17	5	31.2	208	10	US-09-728-721-65
18	5	31.2	216	10	US-09-747-155-47
19	5	31.2	216	10	US-09-747-155-51

20	5	31.2	216	10	US-09-747-155-116
21	5	31.2	220	9	US-10-107-868-25
22	5	31.2	220	9	US-09-989-545-19
23	5	31.2	220	9	US-10-301-056-25
24	5	31.2	220	12	US-10-107-828-25
25	5	31.2	220	12	US-10-107-907-25
26	5	31.2	220	9	US-09-738-626-4341
27	5	31.2	308	9	US-09-738-626-3808
28	5	31.2	312	9	US-09-864-029-2
29	5	31.2	312	9	US-09-864-029-4
30	5	31.2	321	10	US-09-815-242-4990
31	5	31.2	321	10	US-09-815-242-10635
32	5	31.2	322	10	US-09-816-028A-48
33	5	31.2	345	10	US-09-864-029-8
34	5	31.2	345	10	US-09-886-055-293
35	5	31.2	388	10	US-09-880-137-5
36	5	31.2	388	10	US-09-880-137-6
37	5	31.2	403	9	US-09-738-626-5328
38	5	31.2	403	10	US-09-852-053-4
39	5	31.2	411	10	US-09-815-242-5128
40	5	31.2	430	9	US-10-058-636-2
41	5	31.2	445	9	US-09-992-598-177
42	5	31.2	445	9	US-09-989-293A-177
43	5	31.2	445	9	US-10-083-547-32
44	5	31.2	445	9	US-09-989-735-177
45	5	31.2	445	9	US-09-990-444-177
46	5	31.2	445	9	US-09-989-730-177
47	5	31.2	445	9	US-09-990-436-177
48	5	31.2	445	9	US-09-991-181-177
49	5	31.2	445	9	US-09-983-687-177
50	5	31.2	445	9	US-09-989-734-177
51	5	31.2	445	9	US-09-997-653-177
52	5	31.2	445	9	US-10-174-590-148
53	5	31.2	445	9	US-10-176-758-148
54	5	31.2	445	9	US-10-063-616-32
55	5	31.2	445	9	US-10-175-737-148
56	5	31.2	445	9	US-09-993-667-177
57	5	31.2	445	9	US-10-063-502-32
58	5	31.2	445	9	US-10-173-706-148
59	5	31.2	445	9	US-10-175-738-148
60	5	31.2	445	9	US-10-175-752-148
61	5	31.2	445	9	US-10-176-482-148
62	5	31.2	445	9	US-10-176-757-148
63	5	31.2	445	9	US-10-176-913-148
64	5	31.2	445	9	US-10-180-552-148
65	5	31.2	445	9	US-10-180-557-148

## ALIGNMENTS

RESULT 1  
US-09-854-133-587  
; Sequence 587, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 587  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-854-133-587

Query Match 100.0%; Score 16; DB 9; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-12;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16  
 | | | | | | | | | | | | | | | |  
 Db 1 FOANCGIDFIIFWIFW 16

## RESULT 2

US-09-738-973-587  
 ; Sequence 587, Application US/09738973  
 ; Patent No. US20020110563A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indrias, Carol Yoseph  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Elliot, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14  
 ; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 587  
 ; LENGTH: 16  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-738-973-587

Query Match 100.0%; Score 16; DB 10; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-12;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16  
 | | | | | | | | | | | | | | | |  
 Db 1 FOANCGIDFIIFWIFW 16

## RESULT 3

US-09-854-133-586  
 ; Sequence 586, Application US/09854133  
 ; Publication No. US20020183499A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Secrist, Heather  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C10  
 ; CURRENT APPLICATION NUMBER: US/09/854,133  
 ; CURRENT FILING DATE: 2001-05-11  
 ; NUMBER OF SEQ ID NOS: 735  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 586  
 ; LENGTH: 97  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-854-133-586

Query Match 100.0%; Score 16; DB 9; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 29;

Best Local Similarity 100.0%; Pred. No. 3.3e-11;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16  
 | | | | | | | | | | | | | | | |  
 Db 35 FOANCGIDFIIFWIFW 50

## RESULT 4

US-09-738-973-586  
 ; Sequence 586, Application US/09738973  
 ; Patent No. US20020110563A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indrias, Carol Yoseph  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Elliot, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14  
 ; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 586  
 ; LENGTH: 97  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-738-973-586

Query Match 100.0%; Score 16; DB 10; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-11;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFWIFW 16  
 | | | | | | | | | | | | | | | |  
 Db 35 FOANCGIDFIIFWIFW 50

## RESULT 5

US-09-820-843A-43  
 ; Sequence 43, Application US/09820843A  
 ; Publication No. US20030039963A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Council of Scientific and Industrial Research  
 ; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE  
 ; FILE REFERENCE: Q63915  
 ; CURRENT APPLICATION NUMBER: US/09/820,843A  
 ; CURRENT FILING DATE: 2001-03-30  
 ; NUMBER OF SEQ ID NOS: 118  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 43  
 ; LENGTH: 41  
 ; TYPE: PRT  
 ; ORGANISM: C. jejuni  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; OTHER INFORMATION: hypothetical protein Cj0344  
 ; NAME/KEY: misc\_feature  
 ; OTHER INFORMATION: gi|6967819  
 US-09-820-843A-43

Query Match 31.2%; Score 5; DB 9; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 29;

```
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12
    |||||
Db 10 DFIIF 14

RESULT 6
US-09-864-761-39223
; Sequence 39223, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39223
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006504.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3

; OTHER INFORMATION: SWISSPROT HIT: O00507, EVALUE 4.00e-03
US-09-864-761-39223

Query Match 31.2%; Score 5; DB 10; Length 44;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
    |||||
Db 9 GIDFI 13

RESULT 7
US-10-073-961-221
; Sequence 221, Application US/10073961
; Publication No. US20030077602A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P113C1
; CURRENT APPLICATION NUMBER: US/10/073,961
; CURRENT FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 09/764,887
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
```

;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/249,299  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/236,327  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/241,785  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/244,617  
;; PRIOR FILING DATE: 2000-11-01  
;; PRIOR APPLICATION NUMBER: 60/225,268  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/236,368  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/251,856  
;; PRIOR FILING DATE: 2000-12-08  
;; PRIOR APPLICATION NUMBER: 60/251,868  
;; PRIOR FILING DATE: 2000-12-08  
;; PRIOR APPLICATION NUMBER: 60/229,344  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/234,997  
;; PRIOR FILING DATE: 2000-09-25  
;; PRIOR APPLICATION NUMBER: 60/229,343  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/229,345  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/229,287  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/229,513  
;; PRIOR FILING DATE: 2000-09-05  
;; PRIOR APPLICATION NUMBER: 60/231,413  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/229,509  
;; PRIOR FILING DATE: 2000-09-05  
;; PRIOR APPLICATION NUMBER: 60/236,367  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/237,039  
;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/237,038  
;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/236,370  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/236,802  
;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/237,037  
;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/237,040  
;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/240,960  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/239,935  
;; PRIOR FILING DATE: 2000-10-13  
;; PRIOR APPLICATION NUMBER: 60/239,937  
;; PRIOR FILING DATE: 2000-10-13  
;; PRIOR APPLICATION NUMBER: 60/241,787  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/246,474  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/246,532  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/249,216  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,210  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/226,681  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/225,759  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/225,213  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/227,182  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/225,214  
;; PRIOR FILING DATE: 2000-08-14

;; PRIOR APPLICATION NUMBER: 60/235,836  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/230,438  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/215,135  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: 60/225,266  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/249,218  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,208  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,213  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,212  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,207  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,245  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,244  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,217  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,211  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,215  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,264  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,214  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,297  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/232,400  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/231,242  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/232,081  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/232,080  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,414  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,244  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/233,064  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/233,063  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,397  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,399  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,401  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/241,808  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,826  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,786  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,221  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/246,475  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/231,243  
;; PRIOR FILING DATE: 2000-09-08

Query Match 31.2%; Score 5; DB 9; Length 67;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 32 DFIIF 36

## RESULT 8

US-09-764-887-221  
; Sequence 221, Application US/09764887  
; Patent No. US20020042096A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P113  
; CURRENT APPLICATION NUMBER: US/09/764,887  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 658  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 221  
; LENGTH: 67  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (8)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (18)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-887-221

Query Match 31.2%; Score 5; DB 10; Length 67;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 32 DFIIF 36

## RESULT 9

US-10-102-806-693  
; Sequence 693, Application US/10102806  
; Publication No. US20030054421A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: P103P1C1  
; CURRENT APPLICATION NUMBER: US/10/102,806  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/925,298  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05881  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 846  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 693  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-102-806-693

Query Match 31.2%; Score 5; DB 9; Length 68;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 47 DFIIF 51

## RESULT 10

US-09-764-891-3247  
; Sequence 3247, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3247  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (70)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-3247

Query Match 31.2%; Score 5; DB 9; Length 82;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QANCG 6  
|||||  
Db 6 QANCG 10

## RESULT 11

US-09-764-891-3077  
; Sequence 3077, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3077  
; LENGTH: 104  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-891-3077

Query Match 31.2%; Score 5; DB 9; Length 104;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFIIF 11  
|||||  
Db 31 IDFIIF 35

## RESULT 12

US-10-047-676A-17  
; Sequence 17, Application US/10047676A  
; Patent No. US20020123105A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Fengxia  
; APPLICANT: Caufield, Page W.  
; APPLICANT: Chen, Ping W.  
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS  
; FILE REFERENCE: UAB-17403/22  
; CURRENT APPLICATION NUMBER: US/10/047,676A  
; CURRENT FILING DATE: 2002-03-21  
; PRIOR APPLICATION NUMBER: US 09/627,376

; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 17  
; LENGTH: 118  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-047-676A-17

Query Match 31.2%; Score 5; DB 12; Length 118;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15  
Db 14 IFWIF 18

RESULT 13  
US-09-738-626-4531  
; Sequence 4531, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4531  
; LENGTH: 166  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4531

Query Match 31.2%; Score 5; DB 9; Length 166;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 141 GIDFI 145

RESULT 14  
US-09-746-660A-94  
; Sequence 94, Application US/09746660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauer, Gregor  
; APPLICANT: Kim, Jun-Won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon

; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; FILE REFERENCE: BGI-121CP2  
; CURRENT APPLICATION NUMBER: US/09/746,660A  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 09/606740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 09/603124  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Vers. 2.0  
; SEQ ID NO 94  
; LENGTH: 166  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-746-660A-94

Query Match 31.2%; Score 5; DB 9; Length 166;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 141 GIDFI 145

RESULT 15  
US-09-746-660A-96  
; Sequence 96, Application US/09746660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauer, Gregor  
; APPLICANT: Kim, Jun-Won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; FILE REFERENCE: BGI-121CP2  
; CURRENT APPLICATION NUMBER: US/09/746,660A  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 09/606740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 09/603124  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Vers. 2.0  
; SEQ ID NO 96  
; LENGTH: 166  
; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum  
US-09-746-660A-96

Query Match 31.2%; Score 5; DB 9; Length 166;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 141 GIDFI 145  
|||||

## RESULT 16

US-09-791-279-200  
; Sequence 200, Application US/09791279  
; Publication No. US20030050456A1

## GENERAL INFORMATION:

; APPLICANT: Vogel, Gabriel

; APPLICANT: Wood, Linda S.

; APPLICANT: Parodi, Luis

; APPLICANT: Lind, Peter

; TITLE OF INVENTION: No. US20030050456A1e1 G Protein-Coupled Receptors

; FILE REFERENCE: 00048.US1

; CURRENT APPLICATION NUMBER: US/09/791.279

; CURRENT FILING DATE: 2001-02-23

; PRIOR APPLICATION NUMBER: 60/184,715

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,725

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,712

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,606

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,602

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,604

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,822

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,710

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,689

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,690

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: 60/184,716

; PRIOR FILING DATE: 2000-02-24

; NUMBER OF SEQ ID NOS: 220

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 200

; LENGTH: 198

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-791-279-200

Query Match 31.2%; Score 5; DB 9; Length 198;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 119 FIIFW 123  
|||||

## RESULT 17

US-09-728-721-65

; Sequence 65, Application US/09728721

; Patent No. US20020061845A1

## GENERAL INFORMATION:

; APPLICANT: Bertin, John

; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THERE

; FILE REFERENCE: 07334-124001

; CURRENT APPLICATION NUMBER: US/09/728,721

; CURRENT FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: 09/340,620  
; PRIOR FILING DATE: 1999-06-28  
; PRIOR APPLICATION NUMBER: US 09/207,359  
; PRIOR FILING DATE: 1998-12-08  
; PRIOR APPLICATION NUMBER: US 09/099,041  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: US 09/019,942  
; PRIOR FILING DATE: 1998-02-06  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 65  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
US-09-728-721-65

Query Match 31.2%; Score 5; DB 10; Length 208;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANC 5  
Db 64 FQANC 68  
|||||

## RESULT 18

US-09-747-155-47

; Sequence 47, Application US/09747155

; Patent No. US20020151692A1

## GENERAL INFORMATION:

; APPLICANT: Rouquier, Sylvie

; APPLICANT: Giorgi, Dominique

; TITLE OF INVENTION: No. US20020151692A1e1 Polypeptides and Nucleic Acids Encod

; FILE REFERENCE: 19904-008 (C009B6834US)

; CURRENT APPLICATION NUMBER: US/09/747,155

; CURRENT FILING DATE: 2000-12-21

; PRIOR APPLICATION NUMBER: 60/171,746

; PRIOR FILING DATE: 1999-12-22

; NUMBER OF SEQ ID NOS: 431

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 47

; LENGTH: 216

; TYPE: PRT

; ORGANISM: Gorilla gorilla

; NAME/KEY: misc\_feature

; LOCATION: (1)..(649)

; OTHER INFORMATION: Taxon = 9593; gene = GGO19; Accession DBJ/EMBL/GenBank =

US-09-747-155-47

Query Match 31.2%; Score 5; DB 10; Length 216;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 83 FIIFW 87  
|||||

## RESULT 19

US-09-747-155-51

; Sequence 51, Application US/09747155

; Patent No. US20020151692A1

## GENERAL INFORMATION:

; APPLICANT: Rouquier, Sylvie

; APPLICANT: Giorgi, Dominique

; TITLE OF INVENTION: No. US20020151692A1e1 Polypeptides and Nucleic Acids Encod

; FILE REFERENCE: 19904-008 (C009B6834US)

; CURRENT APPLICATION NUMBER: US/09/747,155

; CURRENT FILING DATE: 2000-12-21

; PRIOR APPLICATION NUMBER: 60/171,746

; PRIOR FILING DATE: 1999-12-22

; NUMBER OF SEQ ID NOS: 431  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 51  
; LENGTH: 216  
; TYPE: PRT  
; ORGANISM: Gorilla gorilla  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(649)  
; OTHER INFORMATION: Taxon = 9593; gene = GGO3; Accession DDBJ/EMBL/GenBank = AF127848  
US-09-747-153-51

Query Match 31.2%; Score 5; DB 10; Length 216;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|||||  
Db 83 FIIFW 87

RESULT 20  
US-09-747-155-116  
; Sequence 116, Application US/09747155  
; Patent No. US20020151692A1  
; GENERAL INFORMATION:  
; APPLICANT: Rouquier, Sylvie  
; TITLE OF INVENTION: No. US20020151692A1 Polypeptides and Nucleic Acids Encoding Sam  
; FILE REFERENCE: 19004-008 (C009B834US)  
; CURRENT APPLICATION NUMBER: US/09/747,155  
; CURRENT FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/171,746  
; PRIOR FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 431  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 116  
; LENGTH: 216  
; TYPE: PRT  
; ORGANISM: Pongo pygmaeus  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(649)  
; OTHER INFORMATION: Taxon = 9600; gene = PPY49; Accession DDBJ/EMBL/GenBank = AF127848  
US-09-747-155-116

Query Match 31.2%; Score 5; DB 10; Length 216;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|||||  
Db 83 FIIFW 87

RESULT 21  
US-10-107-868-25  
; Sequence 25, Application US/10107868  
; Patent No. US20020156242A1  
; GENERAL INFORMATION:  
; APPLICANT: Tamatani, Takuya  
; APPLICANT: Tezuka, Katsunari  
; TITLE OF INVENTION: CELL SURFACE MOLECULE MEDIATING CELL  
; TITLE OF INVENTION: ADHESION AND SIGNAL TRANSMISSION  
; FILE REFERENCE: 06501-039002  
; CURRENT APPLICATION NUMBER: US/10/107,868  
; CURRENT FILING DATE: 2002-03-26  
; PRIOR APPLICATION NUMBER: 09/561,308  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 09/383,551  
; PRIOR FILING DATE: 1999-08-26  
; PRIOR APPLICATION NUMBER: PCT/JP98/00837  
; PRIOR FILING DATE: 1998-02-27

; PRIOR APPLICATION NUMBER: JAPAN 09-62290  
; PRIOR FILING DATE: 1997-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 10-62217  
; PRIOR FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-868-25

Query Match 31.2%; Score 5; DB 9; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|||||  
Db 174 FIIFW 178

RESULT 22  
US-09-989-545-19  
; Sequence 19, Application US/09989545  
; Patent No. US20020164697A1  
; GENERAL INFORMATION:  
; APPLICANT: Lehar, Sophie  
; APPLICANT: Manning, Stephen  
; APPLICANT: Coyle, Anthony J.  
; APPLICANT: Gutierrez-Ramos, Jose-Carlos  
; TITLE OF INVENTION: No. US20020164697A1 Th2-Specific Molecules and Uses Thereof  
; FILE REFERENCE: 5800-10B  
; CURRENT APPLICATION NUMBER: US/09/989,545  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 09/168,229  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 09/258,670  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-989-545-19

Query Match 31.2%; Score 5; DB 9; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
|||||  
Db 174 FIIFW 178

RESULT 23  
US-10-301-056-25  
; Sequence 25, Application US/10301056  
; Publication No. US20030083472A1  
; GENERAL INFORMATION:  
; APPLICANT: Tamatani, Takuya  
; APPLICANT: Tezuka, Katsunari  
; TITLE OF INVENTION: CELL SURFACE MOLECULE MEDIATING CELL  
; TITLE OF INVENTION: ADHESION AND SIGNAL TRANSMISSION  
; FILE REFERENCE: 06501-039001  
; CURRENT APPLICATION NUMBER: US/10/301,056  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US/09/383,551  
; PRIOR FILING DATE: 1999-08-26  
; PRIOR APPLICATION NUMBER: PCT/JP98/00837  
; PRIOR FILING DATE: 1998-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 09-62290  
; PRIOR FILING DATE: 1997-02-27



; PRIOR APPLICATION NUMBER: JAPAN 10-62217  
; PRIOR FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-301-056-25

Query Match 31.2%; Score 5; DB 9; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 174 FIIFW 178

## RESULT 24

US-10-107-828-25  
; Sequence 25, Application US/10107828  
; Patent No. US20020115831A1  
; GENERAL INFORMATION:  
; APPLICANT: Tamatani, Takuya  
; APPLICANT: Tezuka, Katsunari  
; TITLE OF INVENTION: CELL SURFACE MOLECULE MEDIATING CELL  
; FILE REFERENCE: 06501-039002  
; CURRENT APPLICATION NUMBER: US/10/107,828  
; CURRENT FILING DATE: 2002-03-26  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/00837  
; PRIOR FILING DATE: 1998-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 09-62290  
; PRIOR FILING DATE: 1997-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 10-62217  
; PRIOR FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-828-25

Query Match 31.2%; Score 5; DB 12; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 174 FIIFW 178

## RESULT 25

US-10-107-907-25  
; Sequence 25, Application US/10107907  
; Patent No. US20020151685A1  
; GENERAL INFORMATION:  
; APPLICANT: Tezuka, Katsunari  
; APPLICANT: Tamatani, Takuya  
; TITLE OF INVENTION: CELL SURFACE MOLECULE MEDIATING CELL  
; FILE REFERENCE: 06501-039002  
; CURRENT APPLICATION NUMBER: US/10/107,907  
; CURRENT FILING DATE: 2002-03-26  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/00837  
; PRIOR FILING DATE: 1998-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 09-62290

; PRIOR FILING DATE: 1997-02-27  
; PRIOR APPLICATION NUMBER: JAPAN 10-62217  
; PRIOR FILING DATE: 1998-02-26  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-907-25

Query Match 31.2%; Score 5; DB 12; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 174 FIIFW 178

## RESULT 26

US-09-738-626-4341  
; Sequence 4341, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4341  
; LENGTH: 228  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4341

Query Match 31.2%; Score 5; DB 9; Length 228;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15  
Db 23 IFWIF 27

## RESULT 27

US-09-738-626-3808  
; Sequence 3808, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 3808  
LENGTH: 305  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3808

Query Match 31.2%; Score 5; DB 9; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
Db 254 NCGID 258

RESULT 28  
US-09-864-029-2  
Sequence 2, Application US/09864029  
Publication No. US20030082174A1  
GENERAL INFORMATION:  
APPLICANT: Padigar, Muralidhara  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Majumder, Kumud  
APPLICANT: Tchernev, Velizar T.  
APPLICANT: Grosse, William M.  
APPLICANT: Szekeres Jr., Edward S.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Burgess, Catherine E.  
APPLICANT: Shimkets, Richard A.  
APPLICANT: Taupier Jr., Raymond J.  
APPLICANT: Casman, Stacie J.  
APPLICANT: Gangolli, Esha A.  
APPLICANT: MacDougall, John R.  
APPLICANT: Stone, David J.  
APPLICANT: Smithson, Glenda  
TITLE OF INVENTION: NOVEL GPCR-LIKE PROTEINS AND NUCLEIC ACIDS ENCODING  
TITLE OF INVENTION: SAME  
FILE REFERENCE: 21402-022  
CURRENT APPLICATION NUMBER: US/09/864,029  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: 60/206,757  
PRIOR FILING DATE: 2000-05-24  
PRIOR APPLICATION NUMBER: 60/214,372  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/219,786  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 60/207,020  
PRIOR FILING DATE: 2000-05-25  
PRIOR APPLICATION NUMBER: 60/220,593  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/239,542  
PRIOR FILING DATE: 2000-10-10  
PRIOR APPLICATION NUMBER: 60/256,402  
PRIOR FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: 60/271,645  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 60/274,809

PRIOR FILING DATE: 2001-03-09  
PRIOR APPLICATION NUMBER: 60/275,590  
PRIOR FILING DATE: 2001-03-13  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 312  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-864-029-2

Query Match 31.2%; Score 5; DB 9; Length 312;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 150 FIIFW 154

RESULT 29  
US-09-864-029-4  
Sequence 4, Application US/09864029  
Publication No. US20030082174A1  
GENERAL INFORMATION:  
APPLICANT: Padigar, Muralidhara  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Majumder, Kumud  
APPLICANT: Tchernev, Velizar T.  
APPLICANT: Grosse, William M.  
APPLICANT: Szekeres Jr., Edward S.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Burgess, Catherine E.  
APPLICANT: Shimkets, Richard A.  
APPLICANT: Taupier Jr., Raymond J.  
APPLICANT: Casman, Stacie J.  
APPLICANT: Gangolli, Esha A.  
APPLICANT: MacDougall, John R.  
APPLICANT: Stone, David J.  
APPLICANT: Smithson, Glenda  
TITLE OF INVENTION: NOVEL GPCR-LIKE PROTEINS AND NUCLEIC ACIDS ENCODING  
TITLE OF INVENTION: SAME  
FILE REFERENCE: 21402-022  
CURRENT APPLICATION NUMBER: US/09/864,029  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: 60/206,757  
PRIOR FILING DATE: 2000-05-24  
PRIOR APPLICATION NUMBER: 60/214,372  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/219,786  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 60/207,020  
PRIOR FILING DATE: 2000-05-25  
PRIOR APPLICATION NUMBER: 60/220,593  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/239,542  
PRIOR FILING DATE: 2000-10-10  
PRIOR APPLICATION NUMBER: 60/256,402  
PRIOR FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: 60/271,645  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 60/274,809  
PRIOR FILING DATE: 2001-03-09  
PRIOR APPLICATION NUMBER: 60/275,590  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 312  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-864-029-4

Query Match 31.2%; Score 5; DB 9; Length 312;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0;

QY 9 FIIFW 13  
Db 150 FIIFW 154

RESULT 30  
US-09-815-242-4990  
; Sequence 4990, Application US/09815242  
; Patent No. US20020061569A1

; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4990

; LENGTH: 321

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-815-242-4990

Query Match 31.2%; Score 5; DB 10; Length 321;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0;

QY 9 FIIFW 13  
Db 157 FIIFW 161

RESULT 31

US-09-815-242-10635

; Sequence 10635, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10635

; LENGTH: 321

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-815-242-10635

Query Match 31.2%; Score 5; DB 10; Length 321;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0;

QY 9 FIIFW 13  
Db 157 FIIFW 161

RESULT 32

US-09-816-028A-48

; Sequence 48, Application US/09816028A

; Patent No. US20020042369A1

; GENERAL INFORMATION:

; APPLICANT: Gilbert, Michel

; APPLICANT: Wakarchuk, Warren W.

; APPLICANT: National Research Council of Canada

; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of

; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics

; FILE REFERENCE: 019633-000111US

; CURRENT APPLICATION NUMBER: US/09/816,028A

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US 60/118,213

; PRIOR FILING DATE: 1999-02-01

; PRIOR APPLICATION NUMBER: US 09/495,406

; PRIOR FILING DATE: 2000-01-31

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 48

; LENGTH: 322

; TYPE: PRT

; ORGANISM: Campylobacter jejuni

; FEATURE:

; OTHER INFORMATION: Campylobacter alpha-2,3-sialyltransferase I (Cst-I)

; OTHER INFORMATION: from C. jejuni OH4384

US-09-816-028A-48

Query Match 31.2%; Score 5; DB 10; Length 322;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0;

QY 5 CGIDF 9  
Db 166 CGIDF 170

RESULT 33

US-09-864-029-8

```

; Sequence 8, Application US/09864029
; Publication No. US20030082174A1
; GENERAL INFORMATION:
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Grosse, William M.
; APPLICANT: Szekeres Jr., Edward S.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: MacDougall, John R.
; APPLICANT: Stone, David J.
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: NOVEL GPCR-LIKE PROTEINS AND NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 21402-022
; CURRENT APPLICATION NUMBER: US/09/864,029
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/206,757
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/214,372
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/219,786
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/207,020
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/220,593
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/239,542
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/256,402
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/271,645
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/274,809
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/275,590
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-029-8

```

```

Query Match      31.2%; Score 5; DB 9; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 DFIIF 12
Db 9 DFIIF 13

```

```

RESULT 34
US-09-886-055-293
; Sequence 293, Application US/09886055
; Patent No. US20020132273A1
; GENERAL INFORMATION:
; APPLICANT: STRYER, LUBERT
; APPLICANT: ZOZULYA, SERGEY
; TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
; FILE REFERENCE: 078003-0277150
; CURRENT APPLICATION NUMBER: US/09/886,055
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 60/213,812

```

```

; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 522
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 293
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-055-293

```

```

Query Match      31.2%; Score 5; DB 10; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 8 DFIIF 12
Db 9 DFIIF 13

```

```

RESULT 35
US-09-880-137-5
; Sequence 5, Application US/09880137
; Patent No. US20020031295A1
; GENERAL INFORMATION:
; APPLICANT: Berstein, Gabriel
; TITLE OF INVENTION: METHODS OF ASSAYING FOR G
; FILE REFERENCE: MNI-131
; CURRENT APPLICATION NUMBER: US/09/880,137
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/186,706
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-137-5

```

```

Query Match      31.2%; Score 5; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 5 CGIDF 9
Db 137 CGIDF 141

```

```

RESULT 36
US-09-880-137-6
; Sequence 6, Application US/09880137
; Patent No. US20020031295A1
; GENERAL INFORMATION:
; APPLICANT: Berstein, Gabriel
; TITLE OF INVENTION: METHODS OF ASSAYING FOR G
; FILE REFERENCE: MNI-131
; CURRENT APPLICATION NUMBER: US/09/880,137
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/186,706
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-137-6

```

```

Query Match      31.2%; Score 5; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Query Match          31.2%; Score 5; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 32 GIDFI 36

RESULT 39
US-09-815-242-5128
; Sequence 5128, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5128
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5128

Query Match          31.2%; Score 5; DB 10; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FILEW 13
Db 304 FILEW 308

RESULT 40
US-10-058-636-2
; Sequence 2, Application US/10058636
; Publication No. US20030049270A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; TITLE OF INVENTION: Lipopolysaccharide alpha-2,3 sialyltransferase of
; FILE REFERENCE: 014137-013210US
; CURRENT APPLICATION NUMBER: US/10/058,636
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US/09/272,960

```

```

QY 5 CGIDF 9
Db 137 CGIDF 141

RESULT 37
US-09-738-626-5328
; Sequence 5328, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOCUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5328
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5328

Query Match          31.2%; Score 5; DB 9; Length 403;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 32 GIDFI 36

RESULT 38
US-09-852-053-4
; Sequence 4, Application US/09852053
; Patent No. US20020055141A1
; GENERAL INFORMATION:
; APPLICANT: BERENS, STEPHAN
; APPLICANT: KALINOWSKI, JORN
; APPLICANT: PUHLER, ALFRED
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM STRAIN WITH
; TITLE OF INVENTION: ENHANCED SECRETION ACTIVITY
; FILE REFERENCE: MAS/21123/280248
; CURRENT APPLICATION NUMBER: US/09/852,053
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: EPO 00110021.3
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; NAME/KEY: PROPEP
; LOCATION: (1)..(403)
; OTHER INFORMATION: secf
US-09-852-053-4

```

; PRIOR FILING DATE: 1999-03-18  
; PRIOR APPLICATION NUMBER: US 60/078,891  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: US 09/272,960  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Campylobacter jejuni  
US-10-058-636-2

Query Match 31.2% Score 5; DB 9; Length 430;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9  
Db 166 CGIDF 170

## RESULT 41

US-09-992-598-177  
; Sequence 177, Application US/09992598  
; Patent No. US20020160384A1

## GENERAL INFORMATION:

; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Guiney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: P2730PIC20  
; CURRENT APPLICATION NUMBER: US/09/992,598  
; CURRENT FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/075945  
; PRIOR FILING DATE: 1998-02-25  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/084600

; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/087106  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087827  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/088021  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088025  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088026  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088028  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088030  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088033  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088326  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088167  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088202  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088212  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088217  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088655  
; PRIOR FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: 60/088734  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088738  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088742  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088810  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088858  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088861  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088876  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089440  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089512  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089532  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089598  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089599  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089600  
; PRIOR FILING DATE: 1998-06-17

; PRIOR APPLICATION NUMBER: 60/089653  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089801  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089907  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089908  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089947  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/089948  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/089952  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/090246  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090252  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090254  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090349  
; PRIOR FILING DATE: 1998-06-23  
; PRIOR APPLICATION NUMBER: 60/090355  
; PRIOR FILING DATE: 1998-06-23  
; PRIOR APPLICATION NUMBER: 60/090429  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090431  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090435  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090444  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090445  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090472  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090535  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090540  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090542  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090557  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090676  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090678  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090690  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090694  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090695  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090696  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090862  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/090863  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/091360  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091478  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091544  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091519  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091626  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091633  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091978

Query Match 31.2%; Score 5; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DEIIF 12  
Db 325 DEIIF 329  
|||||

RESULT 42  
US-09-989-293A-177  
; Sequence 177, Application US/09989293A  
; Patent No. US20020177164A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: P2730P1C66  
; CURRENT APPLICATION NUMBER: US/09/989,293A  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/075945  
; PRIOR FILING DATE: 1998-02-25  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/084600  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/087106  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02

✓	PRIOR FILING DATE:	1998-06-18
✓	PRIOR APPLICATION NUMBER:	60/089947
✓	PRIOR FILING DATE:	1998-06-19
✓	PRIOR APPLICATION NUMBER:	60/089948
✓	PRIOR FILING DATE:	1998-06-19
✓	PRIOR APPLICATION NUMBER:	60/089952
✓	PRIOR FILING DATE:	1998-06-19
✓	PRIOR APPLICATION NUMBER:	60/090246
✓	PRIOR FILING DATE:	1998-06-22
✓	PRIOR APPLICATION NUMBER:	60/090252
✓	PRIOR FILING DATE:	1998-06-22
✓	PRIOR APPLICATION NUMBER:	60/090254
✓	PRIOR FILING DATE:	1998-06-22
✓	PRIOR APPLICATION NUMBER:	60/090349
✓	PRIOR FILING DATE:	1998-06-23
✓	PRIOR APPLICATION NUMBER:	60/090355
✓	PRIOR FILING DATE:	1998-06-23
✓	PRIOR APPLICATION NUMBER:	60/090429
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090431
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090435
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090444
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090445
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090472
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090535
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090540
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090542
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090557
✓	PRIOR FILING DATE:	1998-06-24
✓	PRIOR APPLICATION NUMBER:	60/090676
✓	PRIOR FILING DATE:	1998-06-25
✓	PRIOR APPLICATION NUMBER:	60/090678
✓	PRIOR FILING DATE:	1998-06-25
✓	PRIOR APPLICATION NUMBER:	60/090690
✓	PRIOR FILING DATE:	1998-06-25
✓	PRIOR APPLICATION NUMBER:	60/090694
✓	PRIOR FILING DATE:	1998-06-25
✓	PRIOR APPLICATION NUMBER:	60/090695
✓	PRIOR FILING DATE:	1998-06-25
✓	PRIOR APPLICATION NUMBER:	60/090696
✓	PRIOR FILING DATE:	1998-06-25
✓	PRIOR APPLICATION NUMBER:	60/090862
✓	PRIOR FILING DATE:	1998-06-26
✓	PRIOR APPLICATION NUMBER:	60/090863
✓	PRIOR FILING DATE:	1998-06-26
✓	PRIOR APPLICATION NUMBER:	60/091360
✓	PRIOR FILING DATE:	1998-07-01
✓	PRIOR APPLICATION NUMBER:	60/091478
✓	PRIOR FILING DATE:	1998-07-02
✓	PRIOR APPLICATION NUMBER:	60/091544
✓	PRIOR FILING DATE:	1998-07-01
✓	PRIOR APPLICATION NUMBER:	60/091519
✓	PRIOR FILING DATE:	1998-07-02
✓	PRIOR APPLICATION NUMBER:	60/091626
✓	PRIOR FILING DATE:	1998-07-02
✓	PRIOR APPLICATION NUMBER:	60/091633
✓	PRIOR FILING DATE:	1998-07-02
✓	PRIOR APPLICATION NUMBER:	60/091978
✓	PRIOR FILING DATE:	1998-07-07
✓	PRIOR APPLICATION NUMBER:	60/091982
✓	PRIOR FILING DATE:	1998-07-07
✓	PRIOR APPLICATION NUMBER:	60/092182
✓	PRIOR FILING DATE:	1998-07-09

31.28; Score 5; DB 9; Length 445;



Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 325 DFIIF 329

## RESULT 43

US-10-063-547-32  
; Sequence 32, Application US/10063547  
; Publication No. US20020182638A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063,547  
; CURRENT FILING DATE: 2002-05-02  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 32  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-547-32

Query Match 31.2%; Score 5; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
|||||  
Db 325 DFIIF 329

## RESULT 44

US-09-989-735-177  
; Sequence 177, Application US/09989735  
; Publication No. US20020193299A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: P2730P1C61  
; CURRENT APPLICATION NUMBER: US/09/989,735  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/075945  
; PRIOR FILING DATE: 1998-02-25  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/084600  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/087106  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087827  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/088021  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088025  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088026  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088028  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088030  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088033  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088326  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088167  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088202  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088212  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088217  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088655  
; PRIOR FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: 60/088734  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088738  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088742  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088810  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088858  
; PRIOR FILING DATE: 1998-06-11

;; PRIOR APPLICATION NUMBER: 60/088861  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088876  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/089105  
;; PRIOR FILING DATE: 1998-06-12  
;; PRIOR APPLICATION NUMBER: 60/089440  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089512  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089514  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089532  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089538  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089598  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089599  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089600  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089653  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089801  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089907  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089908  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089947  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/089948  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/089952  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/090246  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090252  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090254  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090349  
;; PRIOR FILING DATE: 1998-06-23  
;; PRIOR APPLICATION NUMBER: 60/090355  
;; PRIOR FILING DATE: 1998-06-23  
;; PRIOR APPLICATION NUMBER: 60/090429  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090431  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090435  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090444  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090445  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090472  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090535  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090540  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090542  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090676  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090678  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090690  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090694

;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090695  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090696  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090862  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/090863  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/091360  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091478  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091544  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091519  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091626  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091633  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091978  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/091982  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/092182  
;; PRIOR FILING DATE: 1998-07-09

Query Match 31.2%; Score 5; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 DFIIF 12  
| | | | |  
Db 325 DFIIF 329

## RESULT 45

US-09-990-444-177  
; Sequence 177, Application US/09990444  
; Publication No. US20020193300A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2730P1C19  
; CURRENT APPLICATION NUMBER: US/09/990,444  
; CURRENT FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16

; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/075945  
; PRIOR FILING DATE: 1998-02-25  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/084600  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/087106  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087827  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/088021  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088025  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088026  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088028  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088030  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088033  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088326  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088167  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088202  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088212  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088217  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088655  
; PRIOR FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: 60/088734  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088738  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088742  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088810  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088858  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088861  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088876  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089440

; PRIOR APPLICATION NUMBER: 60/090863  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/091360  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091478  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091544  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091519  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091626  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091633  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091978  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/091982  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/092182  
; PRIOR FILING DATE: 1998-07-09

Query Match 31.2%; Score 5; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 DFIIF 12  
|||||  
Db 325 DFIIF 329

Search completed: May 11, 2003, 20:21:14  
Job time : 6.23894 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:07:52 ; Search time 4.95575 Seconds

(without alignments)  
310.377 Million cell updates/sec

Title: US-09-854-133-587

Perfect score: 16

Sequence: 1 FQANCGIDFIWFIFW 16

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database :

PIR\_73:\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	43.8	483	S41689	cytochrome-c oxida
2	6	37.5	156	G89233	N-terminal acetyl
3	6	37.5	205	E64621	conserved hypotet
4	6	37.5	283	E64361	nicotinate-nucleot
5	6	37.5	330	T25939	hypothetical prote
6	6	37.5	360	T35783	probable secreted
7	6	37.5	372	JQ2135	NADH2 dehydrogenas
8	6	37.5	430	T46216	hypothetical prote
9	6	37.5	590	1 QOUTC5	NADH2 dehydrogenas
10	6	37.5	679	1 QOCV5	trifunctional enzy
11	6	37.5	1091	D F83928	hypothetical prote
12	5	31.2	41	G81376	hypothetical prote
13	5	31.2	106	S76060	hypothetical prote
14	5	31.2	120	D89910	large-conductance
15	5	31.2	120	AH1280	hypothetical prote
16	5	31.2	122	G97899	hypothetical prote
17	5	31.2	128	1 S27723	cytochrome cytm pr
18	5	31.2	128	A69214	hypothetical prote
19	5	31.2	128	G64155	vhcC protein - Hae
20	5	31.2	135	B31933	Ig heavy chain pre
21	5	31.2	136	T16723	hypothetical prote
22	5	31.2	138	H97024	uncharacterized sm
23	5	31.2	160	B96986	probable tRNA-meth
24	5	31.2	163	AF2342	hypothetical prote
25	5	31.2	173	B89885	conserved hypotet
26	5	31.2	180	B98134	hypothetical prote
27	5	31.2	180	AH3153	hypothetical prote
28	5	31.2	181	G82911	hypothetical prote
29	5	31.2	187	E81383	probable integral

30	5	31.2	195	2	H86768	GTP-binding protei
31	5	31.2	195	2	A48957	orf 2 5' of pepC -
32	5	31.2	205	2	JN0755	urease accessory p
33	5	31.2	208	2	AC1787	thymidylate kinase
34	5	31.2	208	2	AD1411	thymidylate kinase
35	5	31.2	210	2	AC2982	RhtB family transp
36	5	31.2	214	2	T22896	hypothetical prote
37	5	31.2	214	2	T22892	hypothetical prote
38	5	31.2	219	2	D81792	conserved hypotet
39	5	31.2	220	1	RWH028	T-cell surface gly
40	5	31.2	220	2	B90393	hypothetical prote
41	5	31.2	225	1	C64201	ribosomal protein
42	5	31.2	225	2	AC0587	KDP operon transcr
43	5	31.2	230	2	H97331	response regulator
44	5	31.2	230	2	T32999	hypothetical prote
45	5	31.2	233	2	AF0367	conserved hypotet
46	5	31.2	247	2	C84913	probable pectinest
47	5	31.2	248	2	H70550	probable 3-oxoadip
48	5	31.2	248	2	AF2533	hypothetical prote
49	5	31.2	250	2	E70452	hypothetical prote
50	5	31.2	258	2	S26782	DNA endonuclease I
51	5	31.2	260	2	T10642	carbonic anhydrase
52	5	31.2	264	2	JC7772	aquaporin 10 - hum
53	5	31.2	265	2	E96992	homolog of eukaryo
54	5	31.2	268	2	G83640	tryptophan synthas
55	5	31.2	268	2	F72240	conserved hypotet
56	5	31.2	269	1	JQ2127	tryptophan synthas
57	5	31.2	269	1	A30788	tryptophan synthas
58	5	31.2	272	2	T28737	hypothetical prote
59	5	31.2	276	2	G85074	hypothetical prote
60	5	31.2	277	2	T50805	CCRA-ASSOCIATED FA
61	5	31.2	278	2	H86824	phosphoesterase ho
62	5	31.2	278	2	E98301	hypothetical prote
63	5	31.2	282	2	T33036	hypothetical prote
64	5	31.2	288	1	S73464	ribosomal protein
65	5	31.2	288	2	S75324	aspartoacylase (EC

# ALIGNMENTS

## RESULT 1

S41689  
cytochrome-c oxidase (EC 1.9.3.1) chain I - Theileria parva mitochondrion (fragme  
C;Species: mitochondrion Theileria parva  
C;Date: 31-Dec-1993 #sequence\_revision 02-Aug-1994 #text\_change 07-Dec-1999  
C;Accession: S41689; S40154  
R;Kairo, A.; Fairlamb, A.H.; Gobright, E.; Nene, V.  
EMBO J. 13, 898-905, 1994  
A;Title: A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA sequ  
A;Reference number: S41689; MUID:94155854; PMID:8112303  
A;Accession: S41689  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-483 <KAI>  
A;Cross-references: EMBL:Z23263; NID:9437862; PIDN:CAA80798.1; PID:9437863  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1993  
C;Genetics:  
C;Genome: mitochondrion  
A;Genetic code: SGC6  
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homolog  
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; ment  
transmembrane protein  
F;18-466/Domain: cytochrome-c oxidase chain I homology <COI>  
F;71,386/Binding site: heme a iron (His) (axial ligands) #status predicted  
F;248-252/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F;252/Binding site: oxygen (Tyr) #status predicted  
F;384/Binding site: heme a3 iron (His) (axial ligand) #status predicted  
Query Match 43.8%; Score 7; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFIIF 12  
 Db 150 GIDFIIF 156  
 RESULT 2  
 G69233  
 N-terminal acetyltransferase complex, subunit ARD1 - Methanobacterium thermoautotrophicum  
 C:Species: Methanobacterium thermoautotrophicum  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 08-Oct-1999  
 C:Accession: G69233  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
 Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.;  
 K., S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
 A:Reference number: A69000; MUID:98037514; PMID:9371463  
 A:Accession: G69233  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-156 <MTH>  
 A:Cross-references: GB:AE000872; GB:AE000666; NID:g2622082; PIDN:AAB85496.1; PID:g262210  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH999  
 C:Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase rimi  
 Query Match 37.5%; Score 6; DB 2; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 6.2;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 FIIFWI 14  
 Db 53 FIIFWI 58  
 RESULT 3  
 E64621  
 conserved hypothetical protein HP0813 - Helicobacter pylori (strain 26695)  
 C:Species: Helicobacter pylori  
 C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 24-Sep-1999  
 C:Accession: E64621  
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
 Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A:Reference number: A64520; MUID:97394467; PMID:9252185  
 A:Accession: E64621  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-205 <TOM>  
 A:Cross-references: GB:AE000593; GB:AE000511; NID:g2313944; PIDN:AAD07862.1; PID:g231394  
 C:Genetics:  
 A:Start codon: TTG  
 C:Superfamily: glyoxalase  
 Query Match 37.5%; Score 6; DB 2; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 7.7;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 GIDFII 11  
 Db 23 GIDFII 28  
 RESULT 4  
 E64361  
 nicotinate-nucleotide diphosphorylase (carboxylating) (EC 2.4.2.19) - Methanococcus jann  
 N:Alternate names: quinolinate phosphoribosyltransferase  
 C:Species: Methanococcus jannaschii  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 03-Jun-2002

C:Accession: E64361  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.;  
 Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glo  
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.;  
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jann  
 A:Reference number: A64300; MUID:96337999; PMID:8688087  
 A:Accession: E64361  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-283 <BUL>  
 A:Cross-references: GB:U67499; GB:L77117; NID:g1591190; PIDN:AAB98483.1; PID:g15  
 C:Genetics:  
 A:Map position: REV436372-435521  
 C:Superfamily: nicotinate-nucleotide pyrophosphorylase (carboxylating)  
 C:Keywords: glycosyltransferase; pentosyltransferase  
 Query Match 37.5%; Score 6; DB 2; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 CGIDFI 10  
 Db 48 CGIDFI 53  
 RESULT 5  
 T25939  
 hypothetical protein ZC142.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 13-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T25939  
 R:Bradshaw, H.  
 A:Submitted to the EMBL Data Library, July 1996  
 A:Description: The sequence of C. elegans cosmid ZC142.  
 A:Reference number: Z20114  
 A:Accession: T25939  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-330 <BRA>  
 A:Cross-references: EMBL:U64841; PIDN:AAB04845.1; GSPDB:GN000023; CESP:ZC142.1  
 A:Experimental source: strain Bristol N2; clone ZC142  
 C:Genetics:  
 A:Gene: CESP:ZC142.1  
 A:Map position: 5  
 A:Introns: 23/3; 164/3; 213/3  
 Query Match 37.5%; Score 6; DB 2; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 IIFWIF 15  
 Db 133 IIFWIF 138  
 RESULT 6  
 T35783  
 probable secreted protein - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
 C:Accession: T35783  
 R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, July 1998  
 A:Reference number: Z21570  
 A:Accession: T35783  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-360 <SEP>  
 A:Cross-references: EMBL:AL031013; PIDN:CAA19788.1; GSPDB:GN000070; SCOEDB:SC8A6.1  
 A:Experimental source: strain A3(2)  
 C:Genetics:

A:Gene: SCOEDB:SC8A6.16

Query Match 37.5%; Score 6; DB 2; Length 360;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCG 6  
| | | | |  
Db 90 FOANCG 95

# RESULT 7

JO2135

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - Plectonema boryanum

C:Species: Plectonema boryanum

C:Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 03-Jun-2002

C:Accession: JQ2135

R:Takahashi, Y.; Shonai, F.; Fujita, Y.; Kohchi, T.; Ohyama, K.; Matsubara, H.

Plant Cell Physiol. 32, 969-981, 1991

A:Title: Structure of a co-transcribed gene cluster, ndh1-frxB-ndh6-ndh4L, cloned from

A:Reference number: JQ2135

A:Accession: JQ2135

A:Molecule type: DNA

A:Residues: 1-372 <TAK>

A:Cross-references: DDBJ:D01014; NID:g216817; PIDN:BAA00814.1; PID:g216818

A:Experimental source: strain M101

C:Genetics:

A:Gene: ndh1

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C:Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match 37.5%; Score 6; DB 2; Length 372;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFWI 14  
| | | | |  
Db 216 FIIFWI 221

# RESULT 8

T46216

hypothetical protein T8P19.230 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 17-Mar-2000

C:Accession: T46216

R:Choise, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa

submitted to the Protein Sequence Database, December 1999

A:Reference number: Z23008

A:Accession: T46216

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-430 <CHO>

A:Cross-references: EMBL:AL133315

A:Experimental source: cultivar Columbia; BAC clone T8P19

C:Genetics:

A:Map position: 3

A:Introns: 144/3

A:Note: T8P19.230

C:Superfamily: Arabidopsis CER2 protein

Query Match 37.5%; Score 6; DB 2; Length 430;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ANCGID 8  
| | | | |  
Db 103 ANCGID 108

# RESULT 9

QOUC5

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Trypanosoma brucei mitochondrion

C:Species: mitochondrion Trypanosoma brucei

C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 03-Jun-2002

C:Accession: A04519; D22845

R:Hensgens, L.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van B

Nucleic Acids Res. 12, 7327-7344, 1984

A:Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift

ondrial maxi-circle DNA.

A:Reference number: A93537; MUID:85037915; PMID:6093040

A:Accession: A04519

A:Molecule type: DNA

A:Residues: 1-590 <HEN>

A:Cross-references: GB:M94286; NID:g343546

A:Note: this translation is not annotated in GenBank entry TRBKPGEN, release 109.

C:Comment: The DNA sequence is from a segment of the 20-kb maxicircle, which is b

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SCC6

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphoryl

Query Match 37.5%; Score 6; DB 1; Length 590;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWIF 15  
| | | | |  
Db 115 IIFWIF 120

# RESULT 10

QOCV5

trifunctional enzyme - cauliflower mosaic virus

N:Alternate names: ORF5 protein

N:Contains: aspartic proteinase (EC 3.4.23.-); endonuclease; RNA-directed DNA pol

C:Species: cauliflower mosaic virus

C:Date: 31-Oct-1980 #sequence\_revision 31-Oct-1980 #text\_change 23-Jul-1999

C:Accession: D90729; D94613; JN0497; A04159

R:Franch, A.; Guille, H.; Jonard, G.; Richards, K.; Hirth, L.

Cell 21, 285-294, 1980

A:Title: Nucleotide sequence of cauliflower mosaic virus DNA.

A:Reference number: A90799; MUID:81001865; PMID:7407912

A:Accession: D90799

A:Molecule type: DNA

A:Residues: 1-679 <FRA>

A:Cross-references: GB:J02048; NID:g58821; PIDN:CAA23460.1; PID:g58821

A:Experimental source: strain Strasbourg

R:Gardner, R.C.; Howarth, A.J.; Hahn, P.; Brown-Luedi, M.; Shepherd, R.J.; Messin

Nucleic Acids Res. 9, 2871-2888, 1981

A:Title: The complete nucleotide sequence of an infectious clone of cauliflower m

A:Reference number: A93729; MUID:82014878; PMID:6269062

A:Accession: D93729

A:Molecule type: DNA

A:Residues: 1-'N', 3-11, 'I', 13-98, 'K', 100-144, 'T', 146-154, 'I', 156-313, 'I', 315-514, 'A'

A:Experimental source: strain CM1841

submitted to the Nucleic Acid Sequence Database, October 1982

A:Reference number: A94613

A:Accession: D94613

A:Molecule type: DNA

A:Residues: 1-5, 'Q', 7-9, 'I', 11, 'NQ', 12-17, 'I', 19-63, 'I', 65-82, 'N', 84-86, 'R', 88-91

A:Experimental source: strain D/H

R:Chenault, K.D.; Melcher, U.

Gene 123, 255-257, 1993

A:Title: The complete nucleotide sequence of cauliflower mosaic virus isolate BB

A:Reference number: JN0493; MUID:93154593; PMID:8428667

A:Accession: JN0497

A:Molecule type: DNA

A:Residues: 1-56, 'I', 58-94, 'R', 96-98, 'K', 100-185, 'K', 187-313, 'I', 315-514, 'K', 516

A:Experimental source: isolate BRC

A:Note: the authors translated the codon CAC for residue 57 as Val

C:Superfamily: cauliflower mosaic virus trifunctional enzyme

C:Keywords: aspartic proteinase; endonuclease; hydrolase; multifunctional enzyme

Query Match 37.5%; Score 6; DB 1; Length 679;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
 Db 109 GIDFII 114  
 |||||

RESULT 11  
 F83928  
 hypothetical protein BH2230 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: F83928  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: F83928  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1091 <STO>  
 A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05949.1; GSPDB:GN00  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH2230

Query Match 37.5%; Score 6; DB 2; Length 1091;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
 Db 880 GIDFII 885  
 |||||

RESULT 12  
 G81376  
 hypothetical protein Cj0344 [imported] - Campylobacter jejuni (strain NCTC 11168)  
 C:Species: Campylobacter jejuni  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C:Accession: G81376  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell  
 Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
 A:Reference number: A81250; MUID:20150912; PMID:10688204  
 A:Accession: G81376  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-41 <PAR>  
 A:Cross-references: GB:AL111168; NID:g6967817; PIDN:CAB74181.1; PID:g696781  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: Cj0344

Query Match 31.2%; Score 5; DB 2; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
 Db 10 DFIIF 14  
 |||||

RESULT 13  
 S76060  
 hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C:Accession: S76060

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyaj  
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synech  
 s.  
 A:Reference number: S74322; MUID:97061201; PMID:8905231  
 A:Accession: S76060  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-106 <KAN>  
 A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BA010038.1; PID  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1999

Query Match 31.2%; Score 5; DB 2; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
 Db 97 DFIIF 101  
 |||||

RESULT 14  
 D89910  
 large-conductance mechanosensitive channel [imported] - Staphylococcus aureus (s  
 C:Species: Staphylococcus aureus  
 C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: D89910  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, I.  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimi  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: D89910  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-120 <KOR>  
 A:Cross-references: GB:BA000018; PID:g13701145; PIDN:BA042440.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: mscl  
 C:Superfamily: yhcC protein

Query Match 31.2%; Score 5; DB 2; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFII 11  
 Db 71 IDFII 75  
 |||||

RESULT 15  
 AH1280  
 hypothetical protein lmo1648 [imported] - Listeria monocytogenes (strain EGD-e)  
 C:Species: Listeria monocytogenes  
 C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AH1280  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; B  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; F  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Kreitz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam,  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.;  
 A:Title: Comparative genomics of Listeria species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AH1280  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-120 <GLA>  
 A:Cross-references: GB:NC\_003210; PIDN:CAC99726.1; PID:g16411084; GSPDB:GN00177  
 A:Experimental source: strain EGD-e



A;Gene: cytm  
C:Superfamily: cytochrome c6; cytochrome c6 homology

Query Match 31.2%; Score 5; DB 2; Length 128;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 5; Conservative 0; Mismatches 0; Indels

QY 7 IDFI 11  
 Db 83 IDFI 87

# RESULT 20

B31933  
 Ig heavy chain precursor V region (Xlg14) - African clawed frog (fragment)  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 23-Jul-1999  
 C:Accession: B31933  
 R:Schwager, J.; Mikoryak, C.A.; Steiner, L.A.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 2245-2249, 1988  
 A:Title: Amino acid sequence of heavy chain from Xenopus laevis IGM deduced from cDNA sequence  
 A:Reference number: A94192; MUID:88176921; PMID:2451244  
 A:Accession: B31933  
 A:Molecule type: mRNA  
 A:Residues: 1-135 <SCH>

A:Cross-references: GB:J03632; NID:g214329; PIDN:AAA49791.1; PID:g214330  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:32-114/Domain: immunoglobulin homology <IMM>

Query Match 31.2%; Score 5; DB 2; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
 Db 1 DFIIF 5

# RESULT 21

T16723  
 Hypothetical protein R12B2.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T16723  
 R:Miller, N.

Submitted to the EMBL Data Library, June 1994  
 A:Description: The sequence of C. elegans cosmid R12B2.  
 A:Reference number: Z18567  
 A:Accession: T16723  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-136 <MIL>  
 A:Cross-references: EMBL:U00066; NID:g495688; PID:g495691; PIDN:AAA50741.1; CESP:R12B2.3  
 A:Experimental source: strain Bristol N2  
 C:Genetics:  
 A:Gene: CESP:R12B2.3  
 A:Introns: 25/1; 61/3; 86/2; 112/3

Query Match 31.2%; Score 5; DB 2; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
 Db 25 DFIIF 29

# RESULT 22

H97024  
 Uncharacterized small membrane protein, homolog of ykva B. subtilis [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C:Accession: H97024  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: H97024

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-138 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK78987.1; PID:g15023920; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC1011

Query Match 31.2%; Score 5; DB 2; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14  
 Db 117 IIFWI 121

# RESULT 23

B96986  
 Probable tRNA-methylase (SpuU class) [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: B96986  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: B96986  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-160 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK78677.1; PID:g15023579; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC0700  
 C:Superfamily: Chlamydomydia pneumoniae rRNA methylase

Query Match 31.2%; Score 5; DB 2; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
 Db 101 DFIIF 105

# RESULT 24

AF2342  
 Hypothetical protein alr4293 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AF2342  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigoin, J.; Taniguchi, M.; Shimizu, K.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata

et al. DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Nostoc sp. strain PCC 7120  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AF2342  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-163 <KUR>  
 A:Cross-references: GB:BA0000019; PIDN:BA075992.1; PID:g17133429; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr4293

Query Match 31.2%; Score 5; DB 2; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15

Db 58 IFWIF 62  
|||||

## RESULT 25

B9885  
conserved hypothetical protein SA0989 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

A:Accession: B9885  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: B9885

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-173 <KUR>

A:CROSS-references: GB:BA000018; PID:g13700942; PIDN:BA842238.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA0989

Query Match

31.2%; Score 5; DB 2; Length 173;

Best Local Similarity 100.0%; Pred. No. 92;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDPII 11

|||||

Db 3 IDPII 7

## RESULT 26

B98134  
hypothetical protein AGR\_L70 [imported] - Agrobacterium tumefaciens (strain C58, Cereon  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002

A:Accession: B98134

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: B98134

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-180 <KUR>

A:CROSS-references: GB:AE007870; PIDN:AAK88596.1; PID:g15158309; GSPDB:GN00170

C:Genetics:

A:Gene: AGR\_L70

A:Map position: linear chromosome

Query Match

31.2%; Score 5; DB 2; Length 180;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 FWIFW 16

|||||

Db 47 FWIFW 51

## RESULT 27

AH3153  
hypothetical protein Atu4852 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

A:Accession: AH3153

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AH3153

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-180 <KUR>

A:CROSS-references: GB:AE008689; PIDN:AA45646.1; PID:g17743370; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu4852

A:Map position: linear chromosome

Query Match 31.2%; Score 5; DB 2; Length 180;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 FWIFW 16

|||||

Db 47 FWIFW 51

## RESULT 28

G82911

hypothetical protein UU275 [imported] - Ureaplasma urealyticum

C:Species: Ureaplasma urealyticum

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

A:Accession: G82911

R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Caswell, G.

submitted to GenBank, February 2000

A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of

A:Reference number: A82870

A:Accession: G82911

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-181 <GLA>

A:CROSS-references: GB:AE002124; GB:AF222894; NID:g6899241; PIDN:AAF30684.1; GSPD

A:Experimental source: serovar 3; biovar 1

C:Genetics:

A:Gene: UU275

A:Genetic code: GGC

Query Match

31.2%; Score 5; DB 2; Length 181;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14

|||||

Db 102 IIFWI 106

## RESULT 29

E81383

probable integral membrane protein Cj0399 [imported] - Campylobacter jejuni (stra

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

A:Accession: E81383

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.;

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.;

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reve

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: E81383

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-187 <PAR>

A:CROSS-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74235.1; PID:

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0399

Query Match

31.2%; Score 5; DB 2; Length 187;

Best Local Similarity 100.0%; Pred. No. 98;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14  
|||||  
Db 77 IIFWI 81

## RESULT 30

H86768

GTP-binding protein [Imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C:Accession: H86768

R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissbach, J.; Ehrlich

Genome Res. 11, 731-755, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: H86768

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-195 <STO>

A:Cross-references: GB:AE005176; PID:g12724115; PIDN:AAK05250.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: ysxL

C:Superfamily: Bacillus subtilis conserved hypothetical protein ysxC; translation elonga

Query Match 31.2%; Score 5; DB 2; Length 195;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12

|||||

Db 172 DFIIF 176

## RESULT 31

A48957

orf 2 5' of pepC - Lactococcus lactis (fragment)

C:Species: Lactococcus lactis

C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999

C:Accession: A48957

R:Chapot-Chartier, M.P.; Nardil, M.; Chopin, M.C.; Chopin, A.; Gripon, J.C.

Appl. Environ. Microbiol. 59, 330-333, 1993

A:Title: Cloning and sequencing of pepC, a cysteine aminopeptidase gene from Lactococcus

A:Reference number: A48957; MUID:93175873; PMID:8439160

A:Accession: A48957

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-195 <CHA>

A:Cross-references: GB:M86245; NID:g293010; PIDN:AAA74513.1; PID:g293011

A:Experimental source: subsp. cremoris AM2

A:Note: sequence extracted from NCBI backbone (NCBIN:125564, NCBI:P:125565)

Query Match

Best Local Similarity 31.2%; Score 5; DB 2; Length 195;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11

|||||

Db 168 IDFI 172

## RESULT 32

JN0755

urease accessory protein ureg - Proteus mirabilis

N:Contains: probable urease activase (EC 6.3.4.-)

C:Species: Proteus mirabilis

C:Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 02-Feb-2001

C:Accession: JN0755; S32046

R:Sriwanthana, B.; Island, M.D.; Mobley, H.L.T.

Gene 129, 103-106, 1993

A:Title: Sequence of the Proteus mirabilis urease accessory gene ureG.  
A:Reference number: JN0755; MUID:93328109; PMID:8335248

A:Accession: JN0755

A:Molecule type: DNA

A:Residues: 1-205 <SRI>

A:Cross-references: EMBL:221940; NID:g287737; PIDN:CAA79934.1; PID:g287738

C:Comment: This protein may play a chaperonin-like role in the insertion of nickel

C:Genetics:

A:Gene: ureG

C:Superfamily: hydrogenase expression/formaton protein hypB

C:Keywords: GTP binding; ligase; molecular chaperone; nucleotide binding; P-loop

F:14-21/Region: nucleotide-binding motif A (P-loop)

F:151-154/Region: GTP-binding NKXD motif

Query Match 31.2%; Score 5; DB 2; Length 205;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11

|||||

Db 194 IDFI 198

## RESULT 33

AC1787

thymidylate kinase homolog lin2841 [Imported] - Listeria innocua (strain Clp11262

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AC1787

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bl

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsi

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam,

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; We

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AC1787

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-208 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC98067.1; PID:g16415376; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lin2841

C:Superfamily: dTMP kinase

Query Match 31.2%; Score 5; DB 2; Length 208;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10

|||||

Db 31 GIDFI 35

## RESULT 34

AD1411

thymidylate kinase homolog lmo2693 [Imported] - Listeria monocytogenes (strain EGD-

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AD1411

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bl

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsi

Science 294, 849-852, 2001

A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam,

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; We

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1411

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-208 <GLA>  
 A:Cross-references: GB:NC\_003210; PIDN:CAD00906.1; PID:g16412193; GSPDB:GN00177  
 A:Experimental source: strain EGD-e  
 C:Genetics:  
 A:Gene: lmo2693  
 C:Superfamily: dTMP kinase

Query Match 31.2%; Score 5; DB 2; Length 208;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 Db 31 GIDFI 35

RESULT 35  
 AC2982  
 RhtB family transporter rhtB [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
 C:Accession: AC2982  
 R:Wood, D.W.; Satubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AC2982  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-210 <KUR>  
 A:Cross-references: GB:AE008689; PIDN:AAL44273.1; PID:g17741859; GSPDB:GN00187  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: rhtB  
 A:Map position: linear chromosome

Query Match 31.2%; Score 5; DB 2; Length 210;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 Db 159 GIDFI 163

RESULT 36  
 T22896  
 hypothetical protein F58B3.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
 C:Accession: T22896  
 R:Harris, B.  
 submitted to the EMBL Data Library, May 1996  
 A:Reference number: Z19633  
 A:Accession: T22896  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-214 <WIL>  
 A:Cross-references: EMBL:273427; PIDN:CAA97801.1; GSPDB:GN000022; CESP:F58B3.3  
 A:Experimental source: clone F58B3  
 C:Genetics:  
 A:Gene: CESP:F58B3.3  
 A:Map position: 4  
 A:Introns: 68/1  
 C:Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match 31.2%; Score 5; DB 2; Length 214;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 Db 20 GIDFI 24

RESULT 37  
 T22892  
 hypothetical protein F58B3.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
 C:Accession: T22892  
 R:Harris, B.  
 submitted to the EMBL Data Library, May 1996  
 A:Reference number: Z19633  
 A:Accession: T22892  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-214 <WIL>  
 A:Cross-references: EMBL:273427; PIDN:CAA97797.1; GSPDB:GN000022; CESP:F58B3.1  
 A:Experimental source: clone F58B3  
 C:Genetics:  
 A:Gene: CESP:F58B3.1  
 A:Map position: 4  
 A:Introns: 68/1  
 C:Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match 31.2%; Score 5; DB 2; Length 214;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 Db 20 GIDFI 24

RESULT 38  
 DB1792  
 conserved hypothetical protein NMA2192 [imported] - Neisseria meningitidis (stra  
 C:Species: Neisseria meningitidis  
 C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C:Accession: DB1792  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.  
 ; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Raj  
 Nature 404, 502-506, 2000  
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis  
 A:Reference number: AB1775; MUID:20222556; PMID:10761919  
 A:Status: preliminary  
 A:Accession: DB1792  
 A:Molecule type: DNA  
 A:Residues: 1-219 <PAR>  
 A:Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85404.1; PID  
 A:Experimental source: serogroup A, strain Z2491  
 C:Genetics:  
 A:Gene: NMA2192

Query Match 31.2%; Score 5; DB 2; Length 219;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 Db 189 GIDFI 193

RESULT 39  
 RWH28  
 T-cell surface glycoprotein CD28 precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 22-Jun-1999  
 C:Accession: A39983; A45895  
 R:Aruffo, A.; Seed, B.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 8573-8577, 1987

A:Title: Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system  
 A:Reference number: A39983; MUID:88068631; PMID:2825196  
 A:Accession: A39983  
 A:Molecule type: mRNA  
 A:Residues: 1-220 <AR>  
 A:Cross-references: GB:J02988; NID:g338444; PIDN:AAA60581.1; PID:g338445  
 R:Lee, K.P.; Taylor, C.; Petryniak, B.; Turka, L.A.; June, C.H.; Thompson, C.B.  
 J. Immunol. 145, 344-352, 1990  
 A:Title: The genomic organization of the CD28 gene. Implications for the regulation of CD28 gene expression  
 A:Reference number: A45895; MUID:90293482; PMID:2162892  
 A:Accession: A45895  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 35-134 <LEE>  
 A:Cross-references: GB:M37813  
 C:Genetics:  
 A:Gene: GDB:CD28  
 A:Cross-references: GDB:118765; OMIM:186760  
 A:Map position: 2q33-2q34  
 C:Superfamily: T-cell surface glycoprotein CD28; Immunoglobulin homology  
 C:Keywords: glycoprotein; homodimer; T-cell; transmembrane protein  
 F:1-18/Domain: signal sequence status predicted <SIG>  
 F:19-220/Product: T-cell surface glycoprotein CD28 #status predicted <MAT>  
 F:19-152/Domain: extracellular #status predicted <EXT>  
 F:133-114/Domain: immunoglobulin homology <IMM>  
 F:153-179/Domain: transmembrane #status predicted <TM>  
 F:180-220/Domain: intracellular #status predicted <INT>  
 F:71,92,105,129/Binding.site: carbohydrate (Asn) (covalent) #status predicted

Query Match 31.2%; Score 5; DB 1; Length 220;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIF 13  
 |||||  
 DB 174 FIIIF 178

RESULT 40  
 B90393  
 hypothetical protein SSO2231 [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 C:Accession: B90393  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
 arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: B90393  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-220 <KUR>  
 A:Cross-references: GB:AE006641; NID:gl3815533; PIDN:AAK42401.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: SSO2231

Query Match 31.2%; Score 5; DB 2; Length 220;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
 |||||  
 DB 164 DFIIIF 168

RESULT 41  
 C64201  
 ribosomal protein S6 modification protein rimK homolog - Mycoplasma genitalium  
 C:Species: Mycoplasma genitalium  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 07-Dec-1999  
 C:Accession: C64201

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischman  
 M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Meri-  
 , C.A.; Venter, J.C.  
 Science 270, 397-403, 1995  
 A:Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; MUID:96026346; PMID:7569993  
 A:Accession: C64201  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-225 <TIGR>  
 A:Cross-references: GB:U39679; GB:U43967; NID:g1045668; PID:g1045680; TIGR:MG012  
 A:Experimental source: strain G-37  
 C:Genetics:  
 A:Genetic code: SGC3  
 C:Superfamily: ribosomal protein S6 modification protein rimK

Query Match 31.2%; Score 5; DB 1; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9  
 |||||  
 DB 183 CGIDF 187

RESULT 42  
 AC0587  
 KDP operon transcription regulatory protein STY0743 [imported] - Salmonella enter-  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AC0587  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Steve-  
 th, T.; Conner, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.;  
 , S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Steve-  
 th, T.; Conner, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enteric  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AC0587  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-225 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD05165.1; PID:gl6501937; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY0743  
 C:Superfamily: ompR protein; response regulator homology

Query Match 31.2%; Score 5; DB 2; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 DB 60 GIDFI 64

RESULT 43  
 H97331  
 response regulator (CheY-like receiver domain and HTH-type DNA-binding domain) [1  
 C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: H97331  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.  
 ; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacter  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: H97331  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-230 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK81443.1; PID:gl5026610; GSPDB:GN00168

Search completed: May 11, 2003, 20:13:43  
Job time : 7.95575 secs

A:Experimental source: Clostridium acetobutylicum AFCC824  
C:Genetics:  
A:Gene: CAC3517  
C:Superfamily: ompR protein; response regulator homology

Query Match 31.2%; Score 5; DB 2; Length 230;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFI 11  
Db 47 IDFI 51

#### RESULT 44

T32999  
hypothetical protein F17E9.11 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T32999  
R:Woessner, J.  
submitted to the EMBL Data Library, February 1998  
A:Description: The sequence of C. elegans cosmid F17E9.  
A:Reference number: Z21262  
A:Accession: T32999  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-230 <WOR>  
A:Cross-references: EMBL:AF047656; PIDN:AAC05110.1; GSPDB:GN00022; CBSP:F17E9.11  
A:Experimental source: strain Bristol N2; clone F17E9  
C:Genetics:  
A:Gene: CBSP:F17E9.11  
A:Map position: 4  
C:Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match 31.2%; Score 5; DB 2; Length 230;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 20 GIDFI 24

#### RESULT 45

AF0367  
conserved hypothetical protein YP03023 [imported] - Yersinia pestis (strain C092)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AF0367  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AF0367  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-233 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC92265.1; PID:g15980976; GSPDB:GN00175  
C:Genetics:  
A:Gene: YP03023

Query Match 31.2%; Score 5; DB 2; Length 233;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
Db 144 GIDFI 148





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:01:37 ; Search time 3.25664 Seconds  
(without alignments)  
203.775 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FQANGIDFIIFWIFW 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6	37.5	283	1 NADC_METJA	Q57916 methanococ
2	6	37.5	372	1 NU1C_PLEBO	Q00242 plectonema
3	6	37.5	590	1 NU5M_TRYBB	P04540 trypanosoma
4	6	37.5	674	1 POL_CAMVD	P03556 cauliflower
5	6	37.5	679	1 POL_CAMVC	P03555 cauliflower
6	6	37.5	679	1 POL_CAMVE	Q02964 cauliflower
7	6	37.5	679	1 POL_CAMVS	P03554 cauliflower
8	6	37.5	680	1 POL_CAMVN	Q00962 cauliflower
9	5	31.2	91	1 YC19_CYACA	Q9tm45 cyanidium c
10	5	31.2	105	1 C55L_SYNX3	P42351 synchocyst
11	5	31.2	120	1 MSCL_STRAH	O68285 staphylococ
12	5	31.2	128	1 MSCL_HAEIN	P44789 haemophilus
13	5	31.2	135	1 HV02_XENLA	P20957 xenopus lae
14	5	31.2	148	1 AROQ_PASMD	P57903 pasteurellia
15	5	31.2	170	1 HEMX_PROMI	Q51887 proteus mir
16	5	31.2	195	1 ENGB_LACIA	Q9cge5 lactococcus
17	5	31.2	195	1 ENGB_LACIC	Q04731 lactococcus
18	5	31.2	195	1 UREG_PROMI	Q06206 proteus mir
19	5	31.2	205	1 CD28_HUMAN	P10747 homo sapien
20	5	31.2	220	1 SCE2_YEAST	P03878 saccharomyc
21	5	31.2	235	1 SCOA_MYCTU	O06167 mycobacteri
22	5	31.2	248	1 AQPA_HUMAN	Q96ps8 homo sapien
23	5	31.2	264	1 TRPA_PSEAE	P07344 pseudomonas
24	5	31.2	268	1 TRPA_PSEPU	P11081 pseudomonas
25	5	31.2	269	1 TRPA_PSEST	P34816 pseudomonas
26	5	31.2	269	1 TRPA_PSEST	P47258 mycoplasma
27	5	31.2	287	1 Y012_MYCPE	P75097 mycoplasma
28	5	31.2	288	1 Y012_MYCPN	P33172 homo sapien
29	5	31.2	307	1 PPA4_RABIT	P11084 oryctolagus
30	5	31.2	307	1 PPA4_RABIT	Q9zbi5 streptomyce
31	5	31.2	310	1 YS19_STRCO	O29011 archaeoglob
32	5	31.2	320	1 YC57_ARCFU	O07107 enterococ
33	5	31.2	321	1 MRAY_ENTFA	

34	5	31.2	330	1 MCE_CHVP1	Q84424 paramecium
35	5	31.2	337	1 A39B_HUMAN	O43861 homo sapien
36	5	31.2	337	1 YHJD_ECOLI	P37442 escherichia
37	5	31.2	338	1 SYFA_AQUAE	O67087 aquifex aeo
38	5	31.2	341	1 CYS1_OSTOS	P25802 ostertagia
39	5	31.2	342	1 CYS1_HAECO	P19092 haemonchus
40	5	31.2	342	1 CYS2_HAECO	P25793 haemonchus
41	5	31.2	342	1 HENZ_RICPR	Q9zc84 rickettsia
42	5	31.2	342	1 Y762_METJA	Q58172 methanococc
43	5	31.2	343	1 AROB_HELPJ	Q9zmf2 helicobacte
44	5	31.2	343	1 AROB_HELPY	P56081 helicobacte
45	5	31.2	351	1 AROB_CAMJE	Q9pnt2 campylobact
46	5	31.2	354	1 US28_HCMVA	P09704 human cytom
47	5	31.2	356	1 AROB_THETN	Q8rb14 thermoaer
48	5	31.2	362	1 AROB_BACSU	P31102 bacillus su
49	5	31.2	380	1 FT22_METJA	Q58039 methanococc
50	5	31.2	388	1 ARRC_HUMAN	P36575 homo sapien
51	5	31.2	400	1 NIFV_ENTAG	Q52070 enterobacte
52	5	31.2	401	1 T11B_RAT	Q08727 rattus norv
53	5	31.2	404	1 EX7L_FUSNN	Q8ren3 fusobacteri
54	5	31.2	407	1 DEOB_ECOLI	P07651 escherichia
55	5	31.2	412	1 SYV_CHLPN	Q9z812 chlamydia p
56	5	31.2	416	1 CINA_BACSU	P46323 bacillus su
57	5	31.2	419	1 O56A_DROME	Q9v87 drosophila
58	5	31.2	439	1 PPNK_MOUSE	P38058 mus musculu
59	5	31.2	446	1 PPNK_HUMAN	O95544 homo sapien
60	5	31.2	469	1 DNAA_TREDE	O87546 treponema d
61	5	31.2	483	1 ACPA_BACAN	Q44643 bacillus an
62	5	31.2	515	1 CALR_MOUSE	Q60755 mus musculu
63	5	31.2	517	1 FU26_YEAST	P31381 saccharomyc
64	5	31.2	549	1 COX1_TRYBB	P04371 trypanosoma
65	5	31.2	587	1 KPYK_BACST	Q02499 bacillus st

## ALIGNMENTS

RESULT 1				
ID	NADC_METJA	STANDARD;	PRT;	283 AA.
AC	Q57916;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Probable nicotinate-nucleotide pyrophosphorylase [carboxylating]			
DE	(EC 2.4.2.19) (Quinolinate phosphoribosyltransferase			
DE	[decarboxylating]) (QAPRTase).			
GN	NADC OR MJ0493;			
OS	Methanococcus jannaschii.			
OC	Archaea; Euryarchaeota; Methanococci; Methanococcales;			
OC	Methanocaldococcaceae; Methanocaldococcus.			
OX	NCBI_TaxID=2190;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;			
RC	MEDLINE=96337999; PubMed=868087;			
RA	Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,			
RA	Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,			
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,			
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,			
RA	Scott J.L., Geophagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,			
RA	Uttterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,			
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,			
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;			
RT	*Complete genome sequence of the methanogenic archaeon, Methanococcus			
RT	jannaschii*.			
RL	Science 273:1058-1073(1996).			
CC	-!- CATALYTIC ACTIVITY: Nicotinate D-ribonucleotide + diphosphate +			
CC	CO(2) = pyridine-2,3-dicarboxylate + 5-phospho-alpha-D-ribose 1-			
CC	diphosphate.			
CC	-!- PATHWAY: NAD biosynthesis; aspartate to NAMN; third (last) step.			
CC	-!- SIMILARITY: BELONGS TO THE NADC/MODD FAMILY.			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U67499; AAB98483.1; -  
 DR HSSP; O06594; IQPO.  
 DR TIGR; MJ0493; -  
 DR InterPro; IPR004393; NaDC.  
 DR InterPro; IPR002638; ORPase.  
 DR Pfam; PF01729; ORPase\_1.  
 DR Pfam; PF02749; ORPase\_N; 1.  
 DR ProDom; PD003988; ORPase; 1.  
 DR TIGRFAMs; TIGR00078; NaDC; 1.  
 KW Pyridine nucleotide biosynthesis; Transferase; Glycosyltransferase;  
 KW Complete proteome.  
 SQ SEQUENCE 283 AA; 31970 MW; BDD118E9CE1401A8 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CGIDFI 10  
 DB 48 CGIDFI 53  
 |||||

RESULT 2  
 NUC\_PLEBO STANDARD; PRT; 372 AA.  
 ID NUC\_PLEBO  
 AC Q00242;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE NADH-plastoquinone oxidoreductase chain 1 homolog (EC 1.6.5.3).  
 GN NDH1.  
 OS Plectonema boryanum.  
 OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.  
 OX NCBI\_TaxID=1184;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IAM M-101;  
 RA Takahashi Y., Shonai F., Fujita Y., Kohchi T., Ohshima K.,  
 RA Matsubara H.;  
 RT "Structure of a co-transcribed gene cluster, ndh1-frxB-ndh6-ndh4L,  
 RT cloned from the filamentous cyanobacterium Plectonema boryanum.";  
 RL Plant Cell Physiol. 32:969-981(1991).  
 CC -1- CATALYTIC ACTIVITY: NADH + plastoquinone = NAD(+) + plastoquinol.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D01014; BAA00814.1; -  
 DR PIR; JQ2135; JQ2135.  
 DR InterPro; IPR001694; Resp\_NADH\_dhl.  
 DR Pfam; PF00146; NADHdh; 1.  
 DR PROSITE; PS00667; COMPLEX1\_NDL\_1; FALSE\_NEG.  
 DR PROSITE; PS00668; COMPLEX1\_NDL\_2; 1.  
 KW Oxidoreductase; NAD; Plastoquinone; Transmembrane.  
 SQ SEQUENCE 372 AA; 40239 MW; 971729ED98F2094D CRC64;

Query Match 37.5%; Score 6; DB 1; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 7.1;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 FIIFWI 14  
 DB 216 FIIFWI 221  
 |||||

RESULT 3  
 NUSM\_TRYBB STANDARD; PRT; 590 AA.  
 ID NUSM\_TRYBB  
 AC P04540;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).  
 GN ND5.  
 OS Trypanosoma brucei brucei.  
 OG Mitochondrion.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI\_TaxID=5702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=85037915; PubMed=6093040;  
 RA Hensgens L.A.M., Brakenhoff J., de Vries B.F., Sloof P., Tromp M.C.,  
 RA van Boom J.H., Benne R.;  
 RT "The sequence of the gene for cytochrome c oxidase subunit I, a  
 RT frameshift containing gene for cytochrome c oxidase subunit II and  
 RT seven unassigned reading frames in Trypanosoma brucei mitochondrion  
 RT maxi-circle DNA.";  
 RL Nucleic Acids Res. 12:7327-7344(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=164;  
 RX MEDLINE=87201680; PubMed=2437452;  
 RA Jasmer D.P., Feagin J.E., Payne M., Stuart K.;  
 RT "Variation of G-rich mitochondrial transcripts among stocks of  
 RT Trypanosoma brucei.";  
 RL Mol. Biochem. Parasitol. 22:259-272(1987).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X01094; CAB57807.1; -  
 DR EMBL; M14820; AAB59225.1; -  
 DR PIR; A04519; QOQTC5.  
 DR InterPro; IPR001750; Oxidored\_q1.  
 DR Pfam; PF00361; Oxidored\_q1; 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.  
 SQ SEQUENCE 590 AA; 71494 MW; CAE231AD9DF5282 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 590;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 IIFWIF 15  
 DB 115 IIFWIF 120  
 |||||

RESULT 4  
 POL\_CAMVD STANDARD; PRT; 674 AA.  
 ID POL\_CAMVD  
 AC P03556;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Enzymatic polypeptide [Contains: Aspartic protease (EC 3.4.23.-);  
 DE Endonuclease; Reverse transcriptase (EC 2.7.7.49)].

```
GN V.
OS Cauliflower mosaic virus (strain D/H) (CamV).
OC Viruses; Retroid viruses; Caulimoviridae; Caulimovirus.
OX NCBI_TaxID=10645;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=83106468; PubMed=7152260;
RA Balazs E., Guilley H., Jonard G., Richards K.;
RT "Nucleotide sequence of DNA from an altered-virulence isolate D/H of
the cauliflower mosaic virus.";
RL Gene 19:239-249(1982).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ [DNA](N).
CC -!- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
CC -!- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; V00140; -; NOT_ANNOTATED_CDS.
CC DR PIR; A04159; OQCV5.
CC DR MEROPS; A03.001; -.
CC DR InterPro; IPR000588; Peptidase_A3.
CC DR InterPro; IPR000477; RVTse.
CC DR Pfam; PF00078; rvt; 1.
CC DR Pfam; PF02160; Peptidase_A3; 1.
CC DR PRINTS; PR00731; CAULIMOPTASE.
CC DR HydroLase; Aspartyl protease; RNA-directed DNA polymerase;
KW Transferase; Endonuclease; Polypeptide.
CC ACT_SITE 45 45
CC FT SIMILAR 40 130 TO RETROVIRUS GAG/POL PROTEASE DOMAIN.
CC FT SIMILAR 260 620 TO RETROVIRUS GAG/POL DNA POLYMERASE
CC FT SIMILAR 260 620 DOMAIN.
CC FT SEQUENCE 679 AA; 78669 MW; E21F8BC528C9DB8D CRC64;
CC
CC Query Match 37.5%; Score 6; DB 1; Length 679;
CC Best Local Similarity 100.0%; Pred. No. 11;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 6 GIDFII 11
CC Db 109 GIDFII 114
CC
CC RESULT 6
CC POL_CAMVE STANDARD; PRT; 679 AA.
CC AC Q02964;
CC DT 01-JUL-1993 (Rel. 26, Created)
CC DT 01-JUL-1993 (Rel. 26, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Enzymatic polypeptide [Contains: Aspartic protease (EC 3.4.23.-);
GN Endonuclease; Reverse transcriptase (EC 2.7.7.49)].
CC OS Cauliflower mosaic virus (strain BBC) (CamV).
CC OC Viruses; Retroid viruses; Caulimoviridae; Caulimovirus.
CC OX NCBI_TaxID=31556;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=93154593; PubMed=8428667;
CC RA Chenault K.D., Melcher U.K.;
CC RT "The complete nucleotide sequence of cauliflower mosaic virus isolate
RT BBC.";
CC RL Gene 123:255-257(1993).
CC CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ [DNA](N).
CC CC -!- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
CC CC -!- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
CC CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
CC CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; M90542; AAA62375.1; -.
CC DR MEROPS; A03.001; -.
CC
```

```

DR InterPro: IPR000588; Peptidase_A3.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00078; rvt; 1.
DR PRINTS: PR00731; CAULIMOPTASE.
KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
FT ACT_SITE 45
FT SIMILAR 40 130
FT SIMILAR 260 620
SQ SEQUENCE 679 AA; 78713 MW; 2D8031B62F1CB05A CRC64;

Query Match
Best Local Similarity 37.5%; Score 6; DB 1; Length 679;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11
DB 109 GIDFII 114

RESULT 7
POL_CAMVS
ID POL_CAMVS STANDARD; PRT; 679 AA.
AC P03554;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Enzymatic polypeptide [Contains: Aspartic protease (EC 3.4.23.-);
DE Endonuclease; Reverse transcriptase (EC 2.7.7.49)].
GN V.
OS Cauliflower mosaic virus (strain Strasbourg) (CaMV).
OC Viruses; Retroviral viruses; Caulimoviridae; Caulimovirus.
OX NCBI_TaxID=10648;
RN [1]
RP SEQUENCE FROM N.A.
RA Frank A., Guillely H., Jonard G., Richards K., Hirth L.;
RT "Nucleotide sequence of cauliflower mosaic virus DNA.";
RL Cell 21:285-294(1980).
CC -|- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N).
CC -|- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
CC -|- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPEPTIDES.
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V00141; CAA23460.1;
DR PIR: A04159; QOCV5.
DR MEROPS: A03.001;
DR InterPro: IPR000588; Peptidase_A3.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00078; rvt; 1.
DR PRINTS: PR00731; CAULIMOPTASE.
KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
FT ACT_SITE 45
FT SIMILAR 40 130
FT SIMILAR 260 620
SQ SEQUENCE 679 AA; 78629 MW; 9EE527BCD460B766 CRC64;

Query Match
Best Local Similarity 37.5%; Score 6; DB 1; Length 679;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11
DB 109 GIDFII 114

RESULT 8
POL_CAMVS
ID POL_CAMVS STANDARD; PRT; 680 AA.
AC Q00962;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Enzymatic polypeptide [Contains: Aspartic protease (EC 3.4.23.-);
DE Endonuclease; Reverse transcriptase (EC 2.7.7.49)].
GN V.
OS Cauliflower mosaic virus (strain NY8153) (CaMV).
OC Viruses; Retroviral viruses; Caulimoviridae; Caulimovirus.
OX NCBI_TaxID=31557;
RN [1]
RP SEQUENCE FROM N.A.
RA Chensault K.D., Steffens D.L., Melcher U.K.;
RT "Nucleotide sequence of cauliflower mosaic virus isolate NY8153.";
RL Plant Physiol. 100:542-545(1992).
CC -|- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N).
CC -|- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
CC -|- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPEPTIDES.
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M90341; AAA46358.1;
DR MEROPS: A03.001;
DR InterPro: IPR000588; Peptidase_A3.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00078; rvt; 1.
DR PRINTS: PR00731; CAULIMOPTASE.
KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
FT ACT_SITE 46
FT SIMILAR 41 131
FT SIMILAR 261 621
SQ SEQUENCE 680 AA; 78665 MW; FCE02E09647C2221 CRC64;

Query Match
Best Local Similarity 37.5%; Score 6; DB 1; Length 680;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11
DB 110 GIDFII 115

RESULT 9
POL_CAMVS
ID POL_CAMVS STANDARD; PRT; 91 AA.
AC Q9TM45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 10.5 kDa protein ycf19.
GN YCF19 OR YCF49.
OS Cyanidium caldarium.

```

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
DB 109 GIDFII 114

#### RESULT 8

POL\_CAMVS  
ID POL\_CAMVS STANDARD; PRT; 680 AA.

AC Q00962;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Enzymatic polypeptide [Contains: Aspartic protease (EC 3.4.23.-);  
DE Endonuclease; Reverse transcriptase (EC 2.7.7.49)].  
GN V.

OS Cauliflower mosaic virus (strain NY8153) (CaMV).  
OC Viruses; Retroviral viruses; Caulimoviridae; Caulimovirus.  
OX NCBI\_TaxID=31557;  
RN [1]

RP SEQUENCE FROM N.A.

RA Chensault K.D., Steffens D.L., Melcher U.K.;  
RT "Nucleotide sequence of cauliflower mosaic virus isolate NY8153.";  
RL Plant Physiol. 100:542-545(1992).

CC -|- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + (DNA)(N).  
CC -|- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.  
CC -|- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPEPTIDES.  
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----

DR EMBL: M90341; AAA46358.1;  
DR MEROPS: A03.001;

DR InterPro: IPR000588; Peptidase\_A3.  
DR InterPro: IPR000477; RVTse.

DR Pfam: PF00078; rvt; 1.  
DR PRINTS: PR00731; CAULIMOPTASE.

KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;  
FT ACT\_SITE 46

FT SIMILAR 41 131  
FT SIMILAR 261 621

SQ SEQUENCE 680 AA; 78665 MW; FCE02E09647C2221 CRC64;  
TO RETROVIRUS GAG/POL PROTEASE DOMAIN.  
TO RETROVIRUS GAG/POL DNA POLYMERASE

Query Match  
Best Local Similarity 37.5%; Score 6; DB 1; Length 680;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
DB 110 GIDFII 115

#### RESULT 9

POL\_CAMVS

ID YC19\_CYACA STANDARD; PRT; 91 AA.

AC Q9TM45;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical 10.5 kDa protein ycf19.  
GN YCF19 OR YCF49.  
OS Cyanidium caldarium.

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

OG Chloroplast.  
 CC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;  
 CC Cyanidium.  
 OX NCBI\_TaxID=2771;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=RK-1;  
 RX MEDLINE=20496959; PubMed=11040290;  
 RX Gloeckner G., Rosenthal A., Valentín K.-U.,  
 RT "The structure and gene repertoire of an ancient red algal plastid  
 genome.";  
 RL J. Mol. Evol. 51:382-390(2000).  
 CC -1- SIMILARITY: BELONGS TO THE YCF19 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 31.2%; Score 5; DB 1; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFTI 11  
 Db 77 IDFTI 81

RESULT 10  
 C55L\_SLYN3 STANDARD; PRT; 105 AA.  
 ID C55L\_SLYN3  
 AC P42351;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytochrome c-553-like precursor.  
 GN CYTM OR SLL1245.  
 OS Synecocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
 OX NCBI\_TaxID=1148;  
 RN SEQUENCE FROM N.A.  
 RA Malakhov M.P., Wada H., Los D.A., Semenenko V.E., Murata N.;  
 RT "A new type of cytochrome c from Synecocystis PCC6803.";  
 RL J. Plant Physiol. 144:259-264(1994).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,  
 RA Okumura S., Shimoto S., Takeuchi C., Wada T., Watanabe A.,  
 RA Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synecocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 CC -1- PTM: BINDS ONE HEME GROUP PER MOLECULE.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D10716; BAA01559.1;  
 DR EMBL; D90912; BAA18172.1; ALT\_INIT.  
 DR HSP; P00080; LHRO.  
 DR InterPro; IPR000345; CytC\_heme\_bind.  
 DR InterPro; IPR000388; Cyt\_C1.  
 DR Pfam; PF00034; cytochrome\_c; 1.  
 DR ProDom; PD004020; Cyt\_C1C; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 KW Electron transport; Photosynthesis; Heme; Signal; Complete proteome.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT CHAIN 30 105 CYTOCHROME C-553-LIKE.  
 FT BINDING 45 45 HEME (COVALENT) (BY SIMILARITY).  
 FT BINDING 48 48 HEME (COVALENT) (BY SIMILARITY).  
 FT METAL 49 49 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT METAL 85 85 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 105 AA; 11458 MW; 21247DEF9D008A94 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANC 5  
 Db 41 FOANC 45

RESULT 11  
 MSCL\_STAAM STANDARD; PRT; 120 AA.  
 ID MSCL\_STAAM  
 AC O68285;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Large-conductance mechanosensitive channel.  
 GN MSCL OR SAV1347 OR SALL182 OR MW1235.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699).  
 OS Staphylococcus aureus (strain N315).  
 OS Staphylococcus aureus (strain MW2), and  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=158878; 158879, 196620, 1280;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Mu50 / ATCC 700699, and N315;  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Iabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 RT aureus.";  
 RL Lancet 357:1225-1240(2001).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=MW2;  
 RX MEDLINE=220404717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.;  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA.";  
 RL Lancet 359:1819-1827(2002).  
 [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=RN450;  
 RX MEDLINE=98294049; PubMed=9632260;  
 Moe P.C., Blount P., Kung C.;

RT RT "Functional and structural conservation in the mechanosensitive  
 RL channel mscL implicates elements crucial for mechanosensation."  
 CC Mol. Microbiol. 28:583-592(1998).  
 CC -1- FUNCTION: CHANNEL THAT OPENS IN RESPONSE TO STRETCH FORCES IN THE  
 CC MEMBRANE LIPID BILAYER. MAY PARTICIPATE IN THE REGULATION OF  
 CC OSMOTIC PRESSURE CHANGES WITHIN THE CELL.  
 CC -1- SUBUNIT: HOMOPENTAMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE MSCL FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AP003362; BAB57509.1;  
 CC DR EMBL: AP003133; BAB42440.1;  
 CC DR EMBL: AP004826; BAB95100.1;  
 CC DR EMBL: AF029731; AAC38560.1;  
 CC DR HSP: O53898; IMSL.  
 CC DR InterPro: IPR001185; MS\_channel.  
 CC DR Pfam: PF01741; MSCL; 1.  
 CC DR PRINTS: PR01264; MECHCHANNEL.  
 CC DR ProDom: PD007253; MS\_channel; 1.  
 CC DR TIGRFAMS: TIGR00220; mscL; 1.  
 CC DR PROSITE: PS01327; MSCL; 1.  
 CC KW Transmembrane; Ionic channel; Complete proteome.  
 CC FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 14 40 BY SIMILARITY.  
 CC FT DOMAIN 41 62 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 63 84 BY SIMILARITY.  
 CC FT DOMAIN 85 120 CYTOPLASMIC (POTENTIAL).  
 CC SQ SEQUENCE 120 AA; 13616 MW; A4DIE6B2A7B7D2E5 CRC64;  
 CC  
 CC Query Match 31.2%; Score 5; DB 1; Length 120;  
 CC Best Local Similarity 100.0%; Pred. No. 39;  
 CC Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 7 IDFII 11  
 CC |||||  
 CC Db 71 IDFII 75

RESULT 12  
 MSCL\_HAEIN STANDARD; PRT; 128 AA.  
 ID HV02\_XENLA  
 AC P44789;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Large-conductance mechanosensitive channel.  
 GN MSCL OR HI0626.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus  
 OC Haemophilus  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.V., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.L., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae

Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: CHANNEL THAT OPENS IN RESPONSE TO STRETCH FORCES IN THE  
 CC MEMBRANE LIPID BILAYER. MAY PARTICIPATE IN THE REGULATION OF  
 CC OSMOTIC PRESSURE CHANGES WITHIN THE CELL (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMOPENTAMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (by similarity).  
 CC -1- SIMILARITY: BELONGS TO THE MSCL FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U32745; AAC22286.1;  
 CC DR HSP: O53898; IMSL.  
 CC DR TIGR: HI0626;  
 CC DR InterPro: IPR001185; MS\_channel.  
 CC DR Pfam: PF01741; MSCL; 1.  
 CC DR PRINTS: PR01264; MECHCHANNEL.  
 CC DR ProDom: PD007253; MS\_channel; 1.  
 CC DR TIGRFAMS: TIGR00220; mscL; 1.  
 CC DR PROSITE: PS01327; MSCL; 1.  
 CC KW Transmembrane; Inner membrane; Ionic channel; Complete proteome.  
 CC FT DOMAIN 1 15 CYTOPLASMIC (BY SIMILARITY).  
 CC FT TRANSMEM 16 42 BY SIMILARITY.  
 CC FT DOMAIN 43 74 PERIPLASMIC (BY SIMILARITY).  
 CC FT TRANSMEM 75 96 BY SIMILARITY.  
 CC FT DOMAIN 97 128 CYTOPLASMIC (BY SIMILARITY).  
 CC SQ SEQUENCE 128 AA; 14198 MW; BD3A68D27200E51A CRC64;  
 CC  
 CC Query Match 31.2%; Score 5; DB 1; Length 128;  
 CC Best Local Similarity 100.0%; Pred. No. 41;  
 CC Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 7 IDFII 11  
 CC |||||  
 CC Db 83 IDFII 87

RESULT 13  
 HV02\_XENLA  
 ID HV02\_XENLA STANDARD; PRT; 135 AA.  
 AC P20957;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig heavy chain V region XIG14 precursor (Fragment).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=88176921; PubMed=2451244;  
 RX Schwager J., Mikoryak C.A., Steiner L.A.;  
 RT "Amino acid sequence of heavy chain from Xenopus laevis IgM deduced  
 RT from cDNA sequence: implications for evolution of immunoglobulin  
 RT domains."  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:2245-2249(1988).  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

```

CC -----
DR EMBL: J03632; AAA49791.1; -.
DR PIR: B31933; B31933.
DR HSSP: P01810; 2FBJ.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
DR Immunoglobulin V region; Signal.
FT SIGNAL 1 18
FT CHAIN 19 135
FT CHAIN 135 135
FT NON_TER 135 AA; 15080 MW; EBC467105C00732E CRC64;
FT SEQUENCE 135 AA; 15080 MW; EBC467105C00732E CRC64;
SQ
Query Match 31.2%; Score 5; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 DFIIF 12
Db 1 DFIIF 5
RESULT 14
AROQ_PASMU STANDARD; PRT; 148 AA.
AC P57903;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 3-dehydroquinolate dehydratase (EC 4.2.1.10) (3-dehydroquinase) (Type II
DE DQase).
GN AROQ OR AROD OR PM1093.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- FUNCTION: Catalyzes a trans-dehydration via an enolate
CC intermediate (By similarity).
CC -1- CATALYTIC ACTIVITY: 3-dehydroquinolate = 3-dehydroshikimate + H(2)O.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC third step.
CC -1- SIMILARITY: BELONGS TO THE TYPE-II 3-DEHYDROQUINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE006150; AAK03177.1; -.
CC HSSP: P36918; 2DHQ.
CC InterPro: IPR001874; Diquinase_II.
CC Pfam: PF01220; Diquinase_II; 1.
CC TIGRFAMS: TIGR01088; aroQ; 1.
CC PROSITE: PS01029; DEHYDROQUINASE_II; 1.
CC Aromatic amino acid biosynthesis; Lyase; Complete proteome.
CC SEQUENCE 148 AA; 16626 MW; DD4FB1CB7EB1934B CRC64;
SQ
Query Match 31.2%; Score 5; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 IDFII 11
Db 71 IDFII 75
RESULT 15
HEMX_PROMI STANDARD; PRT; 170 AA.
ID HEMX_PROMI
AC Q51887;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Putative uroporphyrin-III C-methyltransferase (EC 2.1.1.107) (Urogen
DE III methylase) (Fragment).
DE HEMX.
GN Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97028791; PubMed=8874804;
RA Trotot P., Sismeiro O., Vivares C., Glaser P., Bresson-Roy A.,
RA Danchin A.;
RT "Comparative analysis of the cya locus in enterobacteria and related
RT Gram-negative facultative anaerobes.";
RL Biochimie 78:277-287(1996).
CC -1- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + uroporphyrin III -
CC 2 S-adenosyl-L-homocysteine + sirohydrochlorin.
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF SIROHEME AND COBALAMIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U22969; AAC44327.1; -.
CC Porphyryn biosynthesis; Transferase; Methyltransferase.
KW NON_TER 170 170
SQ SEQUENCE 170 AA; 18243 MW; DCA77C4EF3E2BE2C CRC64;
Query Match 31.2%; Score 5; DB 1; Length 170;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 IDFII 11
Db 120 IDFII 124
RESULT 16
ENGB_LACLA STANDARD; PRT; 195 AA.
ID ENGB_LACLA
AC Q9CGE5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable GTP-binding protein engB.
GN ENGB OR LLI152.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ILL403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarre K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis ILL403.";

```

```

RL Genome Res. 11:731-753(2001).
CC -1- FUNCTION: NECESSARY FOR NORMAL CELL DIVISION AND FOR THE
CC MAINTENANCE OF NORMAL SEPTATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ENGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE006347; AK05250.1;
CC InterPro: IPR005289; GTP-binding_dom.
CC PRINTS: PR0001806; Ras_trnsfrmng.
CC TIGRFAMS: TIGR00650; MG442; 1.
CC Cell division; Septation; GTP-binding; Complete proteome.
CC NP_BIND 32 39 GTP (POTENTIAL).
CC NP_BIND 77 81 GTP (POTENTIAL).
CC NP_BIND 144 147 GTP (POTENTIAL).
CC SEQUENCE 195 AA; 22551 MW; F2750DB86AF93FEF CRC64;
CC -----
CC Query Match 31.2%; Score 5; DB 1; Length 195;
CC Best Local Similarity 100.0%; Pred. No. 55;
CC Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 8 DRIIF 12
CC Db 172 DFIIF 176
CC -----
CC RESULT 17
CC ENGB_LACILC STANDARD; PRT; 195 AA.
CC ID Q916GI;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Probable GTP-binding protein engb.
CC GN ENGB.
CC OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
CC OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
CC OX NCBI_TaxID=1359;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=MGL363;
CC RA Skinner M.M., Trempey J.E.;
CC RT "Sequence analysis and comparison of the clpX region from Lactococcus
CC lactis.";
CC RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC CC -1- FUNCTION: NECESSARY FOR NORMAL CELL DIVISION AND FOR THE
CC MAINTENANCE OF NORMAL SEPTATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ENGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF236863; AA063739.1;
CC InterPro: IPR005289; GTP-binding_dom.
CC PRINTS: PR0001806; Ras_trnsfrmng.
CC TIGRFAMS: TIGR00650; MG442; 1.
CC Cell division; Septation; GTP-binding.
CC NP_BIND 32 39 GTP (POTENTIAL).
CC NP_BIND 77 81 GTP (POTENTIAL).
CC NP_BIND 144 147 GTP (POTENTIAL).

```

```

SQ SEQUENCE 195 AA; 22460 MW; F2751A8499F9250F CRC64;
Query Match 31.2%; Score 5; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 8 DFIIF 12
Db 172 DFIIF 176
-----
RESULT 18
YPEC_LACILC STANDARD; PRT; 195 AA.
ID Q04731;
AC 004731;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in PEPC 5'region (ORF 2) (Fragment).
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AM2;
RX MEDLINE=93175873; PubMed=8439160;
RA Chapot-Chartier M.P., Nardi M., Chopin M.-C., Chopin A., Gripon J.-C.;
RT "Cloning and sequencing of pepC, a cysteine aminopeptidase gene from
RT Lactococcus lactis subsp. cremoris AM2.";
RL Appl. Environ. Microbiol. 59:330-333(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M86245; AAA74513.1;
CC PIR; A48957;
CC InterPro: IPR002819; HD.
CC Pfam: PF01966; HD; 1.
CC KW Hypothetical protein.
CC FT NON_TER
CC FT 1
CC SQ SEQUENCE 195 AA; 22186 MW; 2A4D27E780ABA507 CRC64;
Query Match 31.2%; Score 5; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 7 IDFI 11
Db 168 IDFI 172
-----
RESULT 19
UREG_PROMI STANDARD; PRT; 205 AA.
ID UREG_PROMI
AC Q06206;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Urease accessory protein ureg.
GN UREG.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HI4320;
RX MEDLINE=93328109; PubMed=8335248;

```





CC CC -1- FUNCTION: ENDONUCLEASE INVOLVED IN INTRON HOMING. INTRODUCES A  
CC CC SPECIFIC DOUBLE-STRAND BREAK AT THE JUNCTION OF THE TWO EXONS A4-  
CC CC A5 OF THE COXI GENE AND THUS MEDIATES THE INSERTION OF THE INTRON  
CC CC INTO AN INTRONLESS STRAIN. RECOGNIZES AND CLEAVE THE SEQUENCE  
CC CC 5'-TTTGAATCTTTGGTCCACCTGAAGTATA-3'  
CC CC -1- SIMILARITY: TO OTHER INTRON NATUPASES.  
CC CC -----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC CC use by non-profit institutions as long as its content is in no way  
CC CC modified and this statement is not removed. Usage by and for commercial  
CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC -----  
CC CC EMBL: V00694; CAA24064.1; ALU1\_INIT.  
CC CC PIR: A04511; QXBY34.  
CC CC REBASE: 2616; I-SceII.  
CC CC SGD: S0007264; A14.  
CC CC InterPro: IPR001982; Endonuc\_LAG/HNH.  
CC CC Pfam: PF00961; LAGLIDAG.1; 2.  
CC CC Mitochondrion; Hydrolase; Nuclease; Endonuclease; Intron homing.  
CC CC KW Mitochondrion; Hydrolase; Nuclease; Endonuclease; Intron homing.  
CC CC SQ SEQUENCE 235 AA; 27460 MW; 81A54BC755C2A131 CRC64;  
Query Match 31.2%; Score 5; DB 1; Length 235;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFI 10  
Db 97 GIDFI 101  
|||||  
RESULT 22  
SCOA\_MYCTU  
ID SCOA\_MYCTU STANDARD; PRT; 248 AA.  
AC O06167;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A  
DE (EC 2.8.3.5) (Succinyl CoA:3-oxoacid CoA-transferase) (OXCT A).  
GN SCOA OR RV2504C OR MT2579 OR WTCY07A7.10C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
FN [1]  
PP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC CC -1- CATALYTIC ACTIVITY: Succinyl-CoA + a 3-oxo acid - succinate + a 3-  
CC CC oxo-acyl-CoA.  
CC CC -1- SUBUNIT: HETERODIMER OF A SUBUNIT A AND A SUBUNIT B (BY  
CC CC SIMILARITY).  
CC CC -1- SIMILARITY: BELONGS TO THE 3-OXOACID COA-TRANSFERASE SUBUNIT A  
CC CC FAMILY.  
CC CC -----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC CC use by non-profit institutions as long as its content is in no way  
CC CC modified and this statement is not removed. Usage by and for commercial  
CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC -----  
CC CC EMBL: Z95556; CAB08922.1;  
CC CC TIGR: MT2579; AAK46883.1;  
CC CC TubercuList; RV2504C;  
CC CC InterPro: IPR004165; CoA\_trans.  
CC CC InterPro: IPR004163; CoA\_transf\_1.  
CC CC Pfam: PF01144; CoA\_trans; 1.  
CC CC PROSITE: PS01273; CoA\_TRANSF.1; 1.  
CC CC Transferase; Complete proteome.  
CC CC FT DOMAIN 24 30 COA-BINDING (POTENTIAL).  
CC CC SQ SEQUENCE 248 AA; 26276 MW; 81340193CC48EBD1 CRC64;  
Query Match 31.2%; Score 5; DB 1; Length 248;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 NCGID 8  
Db 53 NCGID 57  
|||||  
RESULT 23  
ID AOPA\_HUMAN  
AC Q36PS8;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Aquaporin 10 (Small intestine aquaporin).  
GN AQP10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
FN [1]  
PP SEQUENCE FROM N.A.  
RC TISSUE=Jejunum;  
RX MEDLINE=21458278; PubMed=11573934;  
RA Hatakeyama S., Yoshida Y., Tani T., Koyama Y., Nihel K., Ohshiro K.,  
RA Kamie J.-I., Yaoita E., Suda T., Hatakeyama K., Yamamoto T.;  
RT "Cloning of a new aquaporin (AQP10) abundantly expressed in duodenum  
RT and jejunum."  
RL Biochem. Biophys. Res. Commun. 287:814-819(2001).  
CC -1- FUNCTION: Forms a water channel. Not permeable to urea and  
CC glycerol. May contribute to water transport in the upper portion  
CC of small intestine.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- TISSUE SPECIFICITY: Expressed exclusively in duodenum and jejunum.  
CC Highest expression in absorptive epithelial cells at the tips of  
CC villi in the jejunum.  
CC -1- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).  
CC -----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC CC use by non-profit institutions as long as its content is in no way  
CC CC modified and this statement is not removed. Usage by and for commercial  
CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC -----

```

CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF159174; AAL25998.1; -.
CC DR Genew; HGNC:16029; AQP10.
CC DR MIM; 606578; -.
CC DR InterPro; IPR000425; MIP_family.
CC DR Pfam; PF00230; MIP; 1.
CC DR ProDom; PD000295; MIP_family; 1.
CC DR PROSITE; PS00221; MIP; 1.
CC KW Transport; Transmembrane.
CC FT DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 28 48 POTENTIAL.
CC FT DOMAIN 49 52 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 53 73 POTENTIAL.
CC FT DOMAIN 74 101 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 102 122 POTENTIAL.
CC FT DOMAIN 123 156 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 157 177 POTENTIAL.
CC FT DOMAIN 178 187 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 188 208 POTENTIAL.
CC FT DOMAIN 209 264 EXTRACELLULAR (POTENTIAL).
CC SQ SEQUENCE 264 AA; 27672 MW; 1332FF9533743412 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ANCGI 7
Db 207 ANCGI 211

RESULT 24
TRPA_PSEAE STANDARD; PRT; 268 AA.
AC P07344;
DT 01-APR-1988 (Rel. 07, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tryptophan synthase alpha chain (EC 4.2.1.20).
GN TRPA OR PA0035.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OC NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=88174343; PubMed=3127651;
RA Hadero A., Crawford I.P.;
RT "Nucleotide sequence of the genes for tryptophan synthase in
RT Pseudomonas aeruginosa.";
RL Mol. Biol. Evol. 3:191-204(1986).
RN [2]
RP REVISIONS.
RA Crawford I.P., Eberly L.;
RL Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.N., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE
CC OF INDOLEGLYCEROL PHOSPHATE TO INDOL AND GLYCERALDEHYDE 3-
CC PHOSPHATE.

```

```

CC -1- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC - L-tryptophan + glyceraldehyde 3-phosphate + H(2)O.
CC -1- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE TRPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M15826; AAA88463.1; -.
CC EMBL; AE004443; AAG03425.1; -.
CC DR PIR; B25355; TSPSAA.
CC DR HSP; P00929; 2WSY.
CC DR InterPro; IPR003009; FMN_enzyme.
CC DR InterPro; IPR002028; Trp_synthaseA.
CC Pfam; PF00290; trp_synthA_1
CC ProDom; PD001535; Trp_synthaseA; 1.
CC DR TIGRfam; TIGR00262; trpA; 1.
CC DR PROSITE; PS00167; TRP_SYNTHASE_ALPHA; 1.
CC KW Tryptophan biosynthesis; Lyase; Complete proteome.
CC FT CONFLICT 42 42 A -> G (IN REF. 1).
CC FT CONFLICT 73 73 D -> E (IN REF. 1).
CC FT CONFLICT 148 148 I -> L (IN REF. 1).
CC FT SEQUENCE 268 AA; 28488 MW; 6EB7052524A9BEAD CRC64;

Query Match 31.2%; Score 5; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 147 GIDFI 151

RESULT 25
TRPA_PSEPU STANDARD; PRT; 269 AA.
AC P11081;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tryptophan synthase alpha chain (EC 4.2.1.20).
GN TRPA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OC NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PPG1 C15;
RX MEDLINE=89335826; PubMed=2503057;
RA Eberly L., Crawford I.P.;
RT "DNA sequence of the tryptophan synthase genes of Pseudomonas
RT putida.";
RL Biochimie 71:521-531(1989).
CC -1- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE
CC OF INDOLEGLYCEROL PHOSPHATE TO INDOL AND GLYCERALDEHYDE 3-
CC PHOSPHATE.
CC -1- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC - L-tryptophan + glyceraldehyde 3-phosphate + H(2)O.
CC -1- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE TRPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X13299; CAA31662.1; -  
 DR PIR; A30768; A30768.  
 DR PIR; S03836; S03836.  
 DR HSSP; P00929; 2WSY.  
 DR InterPro; IPR003009; FMN\_enzyme.  
 DR InterPro; IPR002028; Trp\_synthaseA.  
 DR Pfam; PF00290; trp\_syntA; 1.  
 DR ProDom; PD001535; Trp\_synthaseA; 1.  
 DR TIGRFAMS; TIGR00262; trpA; 1.  
 DR PROSITE; PS00167; TRP\_SYNTHASE\_ALPHA; 1.  
 KW Tryptophan biosynthesis; Lyase.  
 SQ SEQUENCE 269 AA; 28460 MW; 79FE339C965FD6F8 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 Db 147 GIDFI 151

RESULT 26  
 TRPA\_PSESY  
 ID TRPA\_PSESY STANDARD; PRT; 269 AA.  
 AC P34816;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Tryptophan synthase alpha chain (EC 4.2.1.20).  
 GN TRPA.  
 OS Pseudomonas syringae (pv. syringae).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 CC NCBI\_TaxID=321;  
 RN [1]

SEQUENCE FROM N.A.  
 MEDLINE=93138427; PubMed=8423001;  
 Auerbach S., Gao J., Gussin G.N.;  
 "Nucleotide sequences of the trpI, trpB, and trpA genes of  
 Pseudomonas syringae: positive control unique to fluorescent  
 Pseudomonads.";  
 Gene 123:25-32(1993).  
 -1- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE  
 OF INDOLEGLYCEROL PHOSPHATE TO INDOLE AND GLYCERALDEHYDE 3-  
 PHOSPHATE.  
 -1- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate  
 = L-tryptophan + glyceraldehyde 3-phosphate + H(2)O.  
 -1- PATHWAY: Tryptophan biosynthesis; fifth (last) step.  
 -1- SUBUNIT: Tetramer of two alpha and two beta chains (BY  
 SIMILARITY).  
 -1- SIMILARITY: BELONGS TO THE TRPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M95710; AAA26013.1; -  
 DR PIR; JQ2127; JQ2127.  
 DR HSSP; P00929; 2WSY.  
 DR InterPro; IPR003009; FMN\_enzyme.  
 DR InterPro; IPR002028; Trp\_synthaseA.  
 DR Pfam; PF00290; trp\_syntA; 1.

DR ProDom; PD001535; Trp\_synthaseA; 1.  
 DR TIGRFAMS; TIGR00262; trpA; 1.  
 DR PROSITE; PS00167; TRP\_SYNTHASE\_ALPHA; 1.  
 KW Tryptophan biosynthesis; Lyase.  
 SQ SEQUENCE 269 AA; 28439 MW; AD71A35FB49B3785 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 Db 147 GIDFI 151

RESULT 27  
 Y012\_MYCGE  
 ID Y012\_MYCGE STANDARD; PRT; 287 AA.  
 AC P47258;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MG012.  
 GN MG012.  
 OS Mycoplasma genitalium.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 CC NCBI\_TaxID=2097;  
 RN [1]

SEQUENCE FROM N.A.  
 STRAIN=ATCC 33530 / G-37;  
 MEDLINE=96026346; PubMed=7569993;  
 Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
 Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
 Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
 Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
 Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
 Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
 "The minimal gene complement of Mycoplasma genitalium.";  
 Science 270:397-403(1995).  
 -1- SIMILARITY: SOME, TO BACTERIAL RIBOSOMAL PROTEIN S6 MODIFICATION  
 PROTEIN (RIMK).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U39680; AAC71228.1; -  
 DR TIGR; MG012; -  
 DR InterPro; IPR004666; RimK\_fam.  
 DR TIGRFAMS; TIGR00768; rimK\_fam; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 287 AA; 32733 MW; 7A14AF52A5AD7BBD CRC64;

Query Match 31.2%; Score 5; DB 1; Length 287;  
 Best Local Similarity 100.0%; Pred. No. 73;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CGIDF 9  
 Db 245 CGIDF 249

RESULT 28  
 Y012\_MYCPN  
 ID Y012\_MYCPN STANDARD; PRT; 288 AA.  
 AC P75097;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein MG012 homolog (D12\_orf288).  
GN MPN016 OR MP138.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=97105885; PubMed=8948633;  
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
RA Herrmann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
pneumoniae.";  
RL Nucleic Acids Res. 24:4420-4449(1996).  
CC -1- SIMILARITY: SOME, TO BACTERIAL RIBOSOMAL PROTEIN S6 MODIFICATION  
CC PROTEIN (RIMK).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AE000016; AAB95786.1; -;  
DR InterPro: IPR003135; ATP-grasp.  
DR Pfam: PF02222; ATP-grasp; 1.  
DR TIGRFAMs: TIGR00768; rimk\_fam; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 288 AA; 32435 MW; D0215C770C9BDF58 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 288;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 5 CGIDF 9  
DB 245 CGIDF 249  
  
RESULT 29  
PP4\_HUMAN  
ID PP4\_HUMAN STANDARD; PRT; 307 AA.  
AC P33172;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Serine/threonine protein phosphatase 4 (EC 3.1.3.16) (Pp4) (Protein  
phosphatase X) (pp-X).  
GN PP4C OR PP4 OR PPX.  
OS Homo sapiens (Human), and  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606, 10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RX MEDLINE=93129628; PubMed=1336397;  
RA Brewis N.D., Cohen P.T.W.;  
RT "Protein phosphatase X has been highly conserved during mammalian  
evolution.";  
RL Biochim. Biophys. Acta 1171:231-233(1992).  
RN [2]  
RP REVISION TO 75.  
RA Cohen P.T.W.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human, and Mouse;  
RX MEDLINE=99057922; PubMed=9837938;

RA Hu M.C.-T., Tang-Oxley Q., Qiu W.R., Wang Y.-P.,  
RA Mihindukulasuriya K.A., Afshar R., Tan T.-H.;  
RT "Protein phosphatase X interacts with c-Rel and stimulates c-  
Rel/nuclear factor kappaB activity.";  
RL J. Biol. Chem. 273:33561-33565(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human; TISSUE=Placenta;  
RA Strausberg R.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Mouse; STRAIN=129/SvJ;  
RX MEDLINE=21564173; PubMed=11707325;  
RA Hu M.C.-T., Shui J.W., Mihindukulasuriya K.A., Tan T.-H.;  
RT "Genomic structure of the mouse Pp4 gene: a developmentally regulated  
protein phosphatase.";  
RL Gene 278:89-99(2001).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Mouse; TISSUE=Breast tumor;  
RA Strausberg R.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COULD BE INVOLVED IN MICROTUBULE ORGANIZATION.  
CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +  
CC phosphate.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; CENTROSOSES.  
CC -1- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-X  
CC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X70218; CAA49753.1; -;  
DR EMBL: AF097996; AAC96318.1; -;  
DR EMBL: AF088911; AAC96297.1; -;  
DR EMBL: BC001416; AAH01416.1; -;  
DR EMBL: AF378669; AAL35110.1; -;  
DR EMBL: BC001993; AAH01993.1; -;  
DR PIR: S28173; S28173.  
DR HSSP: P08129; LEJW.  
DR Genew; HGNC:9319; Ppp4C.  
DR MIM: 602035; -;  
DR MGD; MGI:1891763; Ppp4C.  
DR InterPro: IPR004843; M-ppetrase.  
DR InterPro: IPR004844; S/T\_phosphatse.  
DR Pfam; PF00149; Metallophos; 1.  
DR PRINTS: PR00114; STPHPTASE.  
DR PRODOM: PD000252; S/T\_phosphatse; 1.  
DR SMART; SM00156; PP2Ac; 1.  
DR PROSITE; PS00125; SER\_THR\_PHOSPHATASE; 1.  
KW Hydrolase; Iron; Manganese.  
FT METAL 54 54 IRON (BY SIMILARITY).  
FT METAL 56 56 IRON (BY SIMILARITY).  
FT METAL 82 82 IRON AND MANGANESE (BY SIMILARITY).  
FT METAL 114 114 MANGANESE (BY SIMILARITY).  
FT ACT\_SITE 115 115 GENERAL ACID (BY SIMILARITY).  
FT METAL 164 164 MANGANESE (BY SIMILARITY).  
FT METAL 238 238 MANGANESE (BY SIMILARITY).  
SQ SEQUENCE 307 AA; 35080 MW; D6FE470A5C6CBAC CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 307;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 8 DFIIF 12  
DB 282 DFIIF 286

```

RESULT 30
PPP4_RABIT
ID PPP4_RABIT STANDARD; PRT; 307 AA.
AC P11084;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Serine/threonine protein phosphatase 4 (EC 3.1.3.16) (Pp4) (Protein
GN phosphatase X) (PP-X).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-New Zealand white; TISSUE=Liver;
RX MEDLINE=90346193; PubMed=2166691;
RA Cohen P.T.W., Brewis N.D., Hughes V., Mann D.J.;
RT "Protein serine/threonine phosphatases; an expanding family.";
RL FEBS Lett. 268:355-359(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-New Zealand white; TISSUE=Liver;
RX MEDLINE=93209245; PubMed=8384557;
RA Brewis N.D., Street A.J., Prescott A.R., Cohen P.T.W.;
RT "PPX, a novel protein serine/threonine phosphatase localized to
RL centrosomes.";
RN EMBO J. 12:987-996(1993).
RN [3]
RP SEQUENCE OF 105-307 FROM N.A.
RC STRAIN-New Zealand white; TISSUE=Liver;
RX MEDLINE=89078593; PubMed=2849555;
RA da Cruz e Silva O.B., da Cruz e Silva E.F., Cohen P.T.W.;
RT "Identification of a novel protein phosphatase catalytic subunit by
RL cDNA cloning.";
RN FEBS Lett. 242:106-110(1988).
CC -!- FUNCTION: COULD BE INVOLVED IN MICROTUBULE ORGANIZATION.
CC -!- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O -> a protein +
CC phosphate.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; CENTROSOMES.
CC -!- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-X
CC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X14031; CAA32191.1; -
DR EMBL; S57412; AAB25913.1; -
DR PIR; S11059; PARBA2.
DR PIR; S02006; S02006.
DR HSP; P08129; IFJM.
DR InterPro; IPR004843; M-ppetrase.
DR InterPro; IPR004844; S/T_phosphatse.
DR Pfam; PF00149; Metallophos; 1.
DR PRINTS; PR00114; STPHPTASE.
DR PRODOM; PD000252; S/T_phosphatse; 1.
DR SMART; SM00156; PP2AC; 1.
DR PROSITE; PS00125; SER_THR_PHOSPHATASE; 1.
DR Hydrolase; Iron; Manganese.
FT METAL 54 54 IRON (BY SIMILARITY).
FT METAL 56 56 IRON (BY SIMILARITY).
FT METAL 82 82 IRON AND MANGANESE (BY SIMILARITY).
FT METAL 114 114 MANGANESE (BY SIMILARITY).
FT ACT_SITE 115 115 GENERAL ACID (BY SIMILARITY).
FT METAL 164 164 MANGANESE (BY SIMILARITY).

```

```

FT METAL 238 238 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 307 AA; 35037 MW; 364A1641F8B22B41 CRC64;
Query Match 31.2%; Score 5; DB 1; Length 307;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 DFIIF 12
DB 282 DFIIF 286
|||||
RESULT 31
YS19_STRCO
ID YS19_STRCO STANDARD; PRT; 310 AA.
AC Q9ZBI5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein SCO6489.
GN SCO6489 OR SC9C7.25.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RN Nature 417:141-147(2002).
CC -!- SIMILARITY: BELONGS TO THE LDCA FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AL035161; CAA22737.1; -
DR InterPro; IPR003507; UPF0094.
DR Pfam; PF02016; UPF0094; 1.
DR Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 310 AA; 32763 MW; 6E2C587F3AE8E615 CRC64;
Query Match 31.2%; Score 5; DB 1; Length 310;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 GIDFI 10
DB 141 GIDFI 145
|||||
RESULT 32
YC57_ARCFU
ID YC57_ARCFU STANDARD; PRT; 320 AA.
AC O29011;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```



RX MEDLINE-97304383; PubMed-9160746;  
 RA Haakansson K., Doherty A.J., Shuman S., Wigley D.B.;  
 RT "X-ray crystallography reveals a large conformational change during  
 RL guanyl transfer by mRNA capping enzymes.";  
 RN Cell 89:545-553(1997).  
 RP [3]  
 RX X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 11-327.  
 RA MEDLINE-98132620; PubMed-9465045;  
 RA Haakansson K., Wigley D.B.;  
 RT "Structure of a complex between a cap analogue and mRNA guanylyl  
 RL transferase demonstrates the structural chemistry of RNA capping.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 95:1505-1510(1998).  
 CC -|- FUNCTION: M-RNA CAPPING. TRANSFERS A GMP CAP ONTO THE END OF MRNA  
 CC THAT TERMINATES WITH A 5'-DIPHOSPHATE TAIL.  
 CC -|- CATALYTIC ACTIVITY: GTP + (5')PP-pur-mRNA - diphosphate +  
 CC G(5')PPP-pur-mRNA.  
 CC -|- COFACTOR: MAGNESIUM OR MANGANESE.  
 CC -|- SUBUNIT: MONOMER.  
 CC -|- SIMILARITY: BELONGS TO THE EUKARYOTIC GTASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U42580; AAC96471.1; -;  
 DR PDB; 1CKM; 07-JUL-97.  
 DR PDB; 1CKN; 07-JUL-97.  
 DR PDB; 1CKO; 28-JAN-98.  
 DR InterPro; IPR001339; mRNA\_cap\_enzyme.  
 DR Pfam; PF01331; mRNA\_cap\_enzyme; 1.  
 KW Transferase; Nucleotidyltransferase; mRNA processing; mRNA capping;  
 KW 3D-structure.  
 FT AC\_SITE 82 82 GUANYLYLATION SITE.  
 FT SEQUENCE 330 AA; 37832 MW; 6AF8A404710812D9 CRC64;  
 Query Match 31.2%; Score 5; DB 1; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 81;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 IDFII 11  
 Db 243 IDFII 247  
 RESULT 35  
 AT9B\_HUMAN STANDARD; PRT; 337 AA.  
 ID AT9B\_HUMAN  
 AC 043861; O60872;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Potential phospholipid-transporting ATPase IIB (EC 3.6.3.1) (HUS5Y-20)  
 DE (fragment).  
 GN ATP9B OR ATP1B OR NEOL1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE-98217376; PubMed-9548971;  
 RA Halleck M.S., Pradhan D., Blackman C.F., Berkes C., Williamson P.L.,  
 RA Schlegel R.A.;  
 RT "Multiple members of a third subfamily of p-type ATPases identified by  
 RL genomic sequences and ESTs.";  
 RN Genome Res. 8:354-361(1998).  
 RP [2]  
 RP SEQUENCE OF 10-337 FROM N.A.

RC TISSUE=Brain;  
 RX MEDLINE-21064499; PubMed-11124703;  
 RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,  
 RA Cannata N., Zimbello R., Lanfranchi G., Valle G.;  
 RT "Characterization of 16 novel human genes showing high similarity to  
 RL yeast sequences.";  
 CC yeast 18:69-80(2001).  
 CC -|- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -|- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (EI-E2  
 CC ATPASES). SUBFAMILY IV.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U78978; AAC05243.1; -;  
 DR EMBL; AJ006268; CAA06934.1; -;  
 DR Genew; HGNC:13541; ATP9B.  
 DR InterPro; IPR001757; ATPase\_EI-E2.  
 DR PROSITE; PS00154; ATPASE\_EI\_E2; PARTIAL.  
 KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;  
 KW Multigene family.  
 FT NON\_TER 1  
 FT DOMAIN 1 120 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 121 141 POTENTIAL.  
 FT DOMAIN 142 153 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 154 172 POTENTIAL.  
 FT DOMAIN 173 202 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 203 221 POTENTIAL.  
 FT DOMAIN 222 228 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 229 251 POTENTIAL.  
 FT DOMAIN 252 257 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 258 278 POTENTIAL.  
 FT DOMAIN 279 295 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 296 320 POTENTIAL.  
 FT DOMAIN 321 337 CYTOPLASMIC (POTENTIAL).  
 FT METAL 64 64 MAGNESIUM (BY SIMILARITY).  
 FT METAL 68 68 MAGNESIUM (BY SIMILARITY).  
 FT CONFLICT 76 76 N -> D (IN REF. 2).  
 FT CONFLICT 196 196 K -> R (IN REF. 2).  
 FT CONFLICT 210 210 I -> S (IN REF. 2).  
 FT CONFLICT 226 226 D -> E (IN REF. 2).  
 FT CONFLICT 263 263 D -> E (IN REF. 2).  
 FT CONFLICT 285 285 K -> R (IN REF. 2).  
 FT CONFLICT 293 293 N -> D (IN REF. 2).  
 FT CONFLICT 321 321 N -> K (IN REF. 2).  
 SQ SEQUENCE 337 AA; 37498 MW; E93C93A44BD826B2 CRC64;  
 Query Match 31.2%; Score 5; DB 1; Length 337;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 ANCGI 7  
 Db 75 ANCGI 79  
 RESULT 36  
 YHJD\_ECOLI STANDARD; PRT; 337 AA.  
 ID YHJD\_ECOLI  
 AC P37642;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yhjD.  
 GN YHJD OR B3522.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;



CC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12 / MG1655;  
RC MEDLINE=94316500; PubMed=8041620;  
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
region from 76.0 to 81.5 minutes";  
RL Nucleic Acids Res. 22:2576-2586(1994).  
CC -1- SIMILARITY: STRONG, TO E.CHRYSANTHEMI HYPOTHETICAL PROTEIN IN  
CC KDGK 5'REGION (AC P45417).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U00039; AAB18498.1; -  
DR EMBL; AE000428; AAC76547.1; -  
DR Ecogene; EGI2248; YHJD.  
DR InterPro; IPR005274; Cons\_hypoth766.  
DR Pfam; PF03631; Ribonuclease\_BN; 1.  
DR TIGRfams; TIGR00766; TIGR00766; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 337 AA; 37911 MW; C41B2A224902E311 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 337;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 12 FWIFW 16  
Db 246 FWIFW 250  
  
RESULT 37  
SYFA\_AQUAE  
ID SYFA\_AQUAE STANDARD; PRT; 338 AA.  
AC 067087;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Phenylalanine-tRNA synthetase alpha chain (EC 6.1.1.20) (Phenylalanine-  
DE -tRNA ligase alpha chain) (PheRS).  
GN PHE5 OR AQ\_953.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
OC Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aufay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus";  
RL Nature 392:353-358(1998).  
CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +  
CC diphosphate + L-phenylalanyl-tRNA(Phe).  
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC PHE-TRNA SYNTHETASE ALPHA CHAIN SUBFAMILY 1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE000715; AAC07051.1; -  
DR HSSP; P27001; IPYS  
DR InterPro; IPR002106; AATRNA\_ligaseII.  
DR InterPro; IPR004529; PheS.  
DR InterPro; IPR004188; Phe\_tRNA\_synth\_N.  
DR InterPro; IPR002319; tRNA\_synth\_2d.  
DR Pfam; PF01409; tRNA-synth\_2d; 1.  
DR Pfam; PF02912; tRNA-synth\_N; 1.  
DR TIGRfams; TIGR00468; pheS; 1.  
DR PROSITE; PS50862; AA\_TRNA\_LIGASE\_II; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
SQ SEQUENCE 338 AA; 39504 MW; DC36592C22FB2305 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 338;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 4 NCGID 8  
Db 293 NCGID 297  
  
RESULT 38  
CYSL\_OSTOS  
ID CYSL\_OSTOS STANDARD; PRT; 341 AA.  
AC P25802;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cathepsin B-like cysteine proteinase 1 precursor (EC 3.4.22.-).  
GN Cp-1  
OS Ostertagia ostertagi.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
OC Trichostrongyloidea; Haemonchidae; Ostertaginae; Ostertagia.  
OX NCBI\_TaxID=6317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Larva;  
RX MEDLINE=93116804; PubMed=1475000;  
RA Pratt D., Bolsvenue R.J., Cox G.N.;  
RT "Isolation of putative cysteine protease genes of Ostertagia  
RT ostertagi";  
RL Mol. Biochem. Parasitol. 56:39-48(1992).  
CC -1- FUNCTION: EXPRESSION OF THE PROTEASE CORRELATES WITH BLOOD-FEEDING  
CC AND SUGGESTS A ROLE FOR THE PROTEASE IN BLOOD DIGESTION.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M88503; AAA29433.1; -  
DR EMBL; M88503; AAA29434.1; ALT\_SEQ.  
DR EMBL; M88504; AAA29435.1; -  
DR HSSP; P07688; 10DQ.  
DR MEROPS; C01.101; -  
DR InterPro; IPR000668; Peptidase\_C1.  
DR InterPro; IPR000169; Shprot\_acsite.  
DR Pfam; PF00112; Peptidase\_C1; 1.  
DR PRINTS; PR00705; PAPA1N.  
DR ProDom; PD000158; peptidase\_C1; 1.

DR PROSITE; PS00139; THIOLEPROTEASE\_CYS; 1.  
 DR PROSITE; PS00639; THIOLEPROTEASE\_HIS; 1.  
 DR PROSITE; PS00640; THIOLEPROTEASE\_ASN; 1.  
 KW Hydrolase; Thiol protease; Zymogen; Glycoprotein; Multigene family;  
 Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT PROPEP 20 88 ACTIVATION PEPTIDE (POTENTIAL).  
 FT CHAIN 89 341 CATHEPSIN B-LIKE CYSTEINE PROTEINASE 1.  
 FT ACT\_SITE 119 119 BY SIMILARITY.  
 FT ACT\_SITE 288 288 BY SIMILARITY.  
 FT ACT\_SITE 308 308 BY SIMILARITY.  
 FT DISULFID 104 133 BY SIMILARITY.  
 FT DISULFID 116 160 BY SIMILARITY.  
 FT DISULFID 152 218 BY SIMILARITY.  
 FT DISULFID 153 156 BY SIMILARITY.  
 FT DISULFID 189 222 BY SIMILARITY.  
 FT DISULFID 197 209 BY SIMILARITY.  
 FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 341 AA; 38439 MW; 07968646E3D920F6 CRC64;  
 Query Match 31.2%; Score 5; DB 1; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 QANCG 6  
 Db 113 QANCG 117  
 RESULT 39  
 CYS1\_HAECO STANDARD; PRT; 342 AA.  
 ID CYS1\_HAECO STANDARD; PRT; 342 AA.  
 AC P19092;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cathepsin B-like cysteine proteinase 1 precursor (EC 3.4.22.-).  
 GN AC-1.  
 OS Haemochus contortus (Barber pole worm).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
 OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
 OX NCBI\_TaxID=6289;  
 RN [J]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Isolate BPL1;  
 RX MEDLINE=90348715; PubMed=2385265;  
 RA Cox G.N., Pratt D., Hageman R., Boisvenue R.J.;  
 RT "Molecular cloning and primary sequence of a cysteine protease  
 expressed by Haemochus contortus adult worms.";  
 RL Mol. Biochem. Parasitol. 41:25-34(1990).  
 CC -[- FUNCTION: EXPRESSION OF THE PROTEASE CORRELATES WITH BLOOD-FEEDING  
 AND SUGGESTS A ROLE FOR THE PROTEASE IN BLOOD DIGESTION.  
 CC -[- DEVELOPMENTAL STAGE: AT LOW LEVEL IN THE THIRD AND FOURTH-STAGE  
 LARVAE, AND ABUNDANT IN ADULT WORMS.  
 CC -[- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M31112; AAA29175.1; -.  
 DR PIR; A45524; A45524.  
 DR HSSP; P07688; IQDQ.  
 DR MEROPS; C01.101; -.  
 DR InterPro; IPR000668; Peptidase\_C1.  
 DR Pfam; PF00112; Peptidase\_C1; 1.  
 DR PRINTS; PR00705; PAPAIN.

DR PRODOM; PD000158; Peptidase\_C1; 1.  
 DR PROSITE; PS00139; THIOLEPROTEASE\_CYS; 1.  
 DR PROSITE; PS00639; THIOLEPROTEASE\_HIS; 1.  
 DR PROSITE; PS00640; THIOLEPROTEASE\_ASN; 1.  
 KW Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal;  
 Multigene family.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT PROPEP 19 86 ACTIVATION PEPTIDE (POTENTIAL).  
 FT CHAIN 87 342 CATHEPSIN B-LIKE CYSTEINE PROTEINASE 1.  
 FT ACT\_SITE 114 114 BY SIMILARITY.  
 FT ACT\_SITE 285 285 BY SIMILARITY.  
 FT ACT\_SITE 305 305 BY SIMILARITY.  
 FT DISULFID 100 128 BY SIMILARITY.  
 FT DISULFID 111 156 BY SIMILARITY.  
 FT DISULFID 147 214 BY SIMILARITY.  
 FT DISULFID 148 152 BY SIMILARITY.  
 FT DISULFID 185 218 BY SIMILARITY.  
 FT DISULFID 193 205 BY SIMILARITY.  
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 342 AA; 38459 MW; D33D62F7419F0471 CRC64;  
 Query Match 31.2%; Score 5; DB 1; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 QANCG 6  
 Db 108 QANCG 112  
 RESULT 40  
 CYS2\_HAECO STANDARD; PRT; 342 AA.  
 ID CYS2\_HAECO STANDARD; PRT; 342 AA.  
 AC P25793;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cathepsin B-like cysteine proteinase 2 precursor (EC 3.4.22.-).  
 GN AC-2.  
 OS Haemochus contortus (Barber pole worm).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
 OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
 OX NCBI\_TaxID=6289;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=91218800; PubMed=2090940;  
 RA Pratt D., Cox G.N., Milhausen M.J., Boisvenue R.J.;  
 RT "A developmentally regulated cysteine protease gene family in  
 Haemochus contortus.";  
 RL Mol. Biochem. Parasitol. 43:181-192(1990).  
 CC -[- FUNCTION: EXPRESSION OF THE PROTEASE CORRELATES WITH BLOOD-FEEDING  
 AND SUGGESTS A ROLE FOR THE PROTEASE IN BLOOD DIGESTION.  
 CC -[- DEVELOPMENTAL STAGE: AT LOW LEVEL IN THE THIRD AND FOURTH-STAGE  
 LARVAE, AND ABUNDANT IN ADULT WORMS.  
 CC -[- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M60213; AAA29171.1; -.  
 DR PIR; A44965; A44965.  
 DR HSSP; P07688; IQDQ.  
 DR MEROPS; C01.101; -.  
 DR InterPro; IPR000668; Peptidase\_C1.

DR InterPro: IPR000169; SHprot\_actsite.  
 DR Pfam: PF00112; Peptidase\_C1; 1.  
 DR PRINTS: PR00705; PAPAIN.  
 DR PRODOM: PD000158; Peptidase\_C1; 1.  
 DR PROSITE: PS00139; THIOLEPROTEASE\_HTS; 1.  
 DR PROSITE: PS00639; THIOLEPROTEASE\_HTS; 1.  
 DR PROSITE: PS00640; THIOLEPROTEASE\_ASN; 1.  
 DR Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal;  
 KW Multigene family.  
 FT SIGNAL 1 18  
 FT PROPEP 19 86  
 FT CHAIN 87 342  
 FT ACT\_SITE 114 114  
 FT ACT\_SITE 285 285  
 FT ACT\_SITE 305 305  
 FT DISULFID 100 128  
 FT DISULFID 111 156  
 FT DISULFID 147 214  
 FT DISULFID 148 152  
 FT DISULFID 185 218  
 FT DISULFID 193 205  
 FT CARBOHYD 99 99  
 FT CARBOHYD 138 138  
 FT CARBOHYD 198 198  
 FT CARBOHYD 296 296  
 SQ SEQUENCE 342 AA; 38406 MW; AF8FC63904903C92 CRC64;  
 Query Match 31.2%; Score 5; DB 1; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 QANCG 6  
 Db 108 QANCG 112  
 RESULT 41  
 ID HEMZ\_RICP STANDARD; PRT; 342 AA.  
 AC Q9ZC84;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ferrochelatase (EC 4.99.1.1) (Prothemin ferro-lyase) (Heme  
 synthetase)  
 DE HEMH OR RP884.  
 GN Rickettsia prowazekii.  
 OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsia;  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Madrid E;  
 RX MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Sichert-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria."  
 RL Nature 396:133-140(1998).  
 CC -1- FUNCTION: CATALYZES THE FERROUS INSERTION INTO PROTOPORPHYRIN IX.  
 CC -1- CATALYTIC ACTIVITY: Protoporphyrin + Fe(2+) = protoheme + 2 H(+).  
 CC -1- PATHWAY: Prothemin biosynthesis; last step.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE FERROCHELATASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AJ235273; CAA15306.1; ALT INIT.  
 DR InterPro: IPR001015; Ferrochelatase.  
 DR Pfam: PF00762; Ferrochelatase; 1.  
 DR PRODOM: PD002792; Ferrochelatase; 1.  
 DR TIGRFAMS; TIGR00109; hemH; 1.  
 DR PROSITE: PS00534; FERROCHELATASE; 1.  
 DR PROSITE: PS00534; FERROCHELATASE; 1.  
 KW Porphyrin biosynthesis; Heme biosynthesis; Lyase; Iron;  
 KW Complete proteome.  
 FT METAL 188 188  
 FT METAL 268 268  
 FT METAL 342 AA; 39610 MW; 33EA8548F5CF0372 CRC64;  
 SQ SEQUENCE 342 AA; 39610 MW; 33EA8548F5CF0372 CRC64;  
 Query Match 31.2%; Score 5; DB 1; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 DFIF 12  
 Db 92 DFIF 96  
 RESULT 42  
 ID Y762\_METJA STANDARD; PRT; 342 AA.  
 AC Q58172;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ0762.  
 GN Methanococcus jannaschii.  
 OS Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii."  
 RL Science 273:1058-1073(1996).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: STRONG, TO M.JANNASCHII MJ0576 AND TO S.POMBE MALATE  
 CC PERMEASE (NAE1).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 242 262 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 304 324 POTENTIAL.
SQ SEQUENCE 342 AA; 39534 MW; 08EFEC3E2C4955D8 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 342;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 93 GIDFI 97

RESULT 43
AROB_HELPJ
ID AROB_HELPJ STANDARD; PRT; 343 AA.
AC Q2MF2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-dehydroquininate synthase (EC 4.2.3.4).
GN AROB OR JHP0268.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- CATALYTIC ACTIVITY: 3-deoxy-arabino-heptulosonate 7-phosphate = 3-
CC dehydroquininate + phosphate.
CC -1- COFACTOR: NAD and a divalent metal cation (By similarity).
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC second step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROQUINATE SYNTHASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: AE001464; AAD05849.1; -
DR HSSP: P07547; IDQS.
DR InterPro: IPR002658; DHQ synthase.
DR Pfam: PF01761; DHQ synthase; 1.
KW Aromatic amino acid biosynthesis; Lyase; NAD; Complete proteome.
SQ SEQUENCE 343 AA; 39106 MW; 21BD1167236ED5B CRC64;

Query Match 31.2%; Score 5; DB 1; Length 343;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 111 GIDFI 115

```

```

RESULT 44
AROB_HELPY
ID AROB_HELPY STANDARD; PRT; 343 AA.
AC P56081;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-dehydroquininate synthase (EC 4.2.3.4).
GN AROB OR HP0283.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- CATALYTIC ACTIVITY: 3-deoxy-arabino-heptulosonate 7-phosphate = 3-
CC dehydroquininate + phosphate.
CC -1- COFACTOR: NAD and a divalent metal cation (By similarity).
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC second step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROQUINATE SYNTHASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: AE000547; AAD07351.1; -
DR HSSP: P07547; IDQS.
DR TIGR: HP0283; -
DR InterPro: IPR002658; DHQ synthase.
DR Pfam: PF01761; DHQ synthase; 1.
KW Aromatic amino acid biosynthesis; Lyase; NAD; Complete proteome.
SQ SEQUENCE 343 AA; 39120 MW; 5AEAC2F4DE816D13 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 343;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10
Db 111 GIDFI 115

RESULT 45
AROB_CAMJE
ID AROB_CAMJE STANDARD; PRT; 351 AA.
AC Q9PNT2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-dehydroquininate synthase (EC 4.2.3.4).
GN AROB OR CJ1008C.
OS Campylobacter jejuni.

```

```

OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OX Campylobacter.
RN NCBI_TaxID=197;
RP [1]
RC SEQUENCE FROM N.A.
RX STRAIN=NCTC 11168;
RA MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Ouail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -1- CATALYTIC ACTIVITY: 3-deoxy-arabino-heptulosonate 7-phosphate = 3-
CC dehydroquininate + phosphate.
CC -1- COFACTOR: NAD and a divalent metal cation (By similarity).
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC second step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROQUINATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL139076; CAB73264.1; -.
DR HSSP; P07547; 1DQS.
DR InterPro; IPR002658; DHQ_synthase.
DR Pfam; PF01761; DHQ_synthase; 1.
KW Aromatic amino acid biosynthesis; Lyase; NAD; Complete proteome.
SQ SEQUENCE 351 AA; 39596 MW; DEF5FF894DE8727 CRC64;
Query Match 31.2%; Score 5; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 GIDFI 10
Db 110 GIDFI 114

```

Search completed: May 11, 2003, 20:12:49  
 Job time : 5.25664 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 19:33:41 ; Search time 8.49557 seconds  
(without alignments)  
388.055 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 16  
Sequence: 1 FOANCGIDFIIFWIFW 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 50.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database :

SPTRMBL.21.\*  
1: sp-archaea.\*  
2: sp-bacteria.\*  
3: sp-fungi.\*  
4: sp-human.\*  
5: sp-invertebrate.\*  
6: sp-mammal.\*  
7: sp-mhc.\*  
8: sp-organella.\*  
9: sp-phage.\*  
10: sp-plant.\*  
11: sp-rodent.\*  
12: sp-virus.\*  
13: sp-vertebrate.\*  
14: sp-unclassified.\*  
15: sp-rvirus.\*  
16: sp-bacteriap.\*  
17: sp-archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	43.8	483	Q36097	Q36097 theileria p
2	6	37.5	156	17	O27080 methanobact
3	6	37.5	205	16	O25493 helicobacte
4	6	37.5	250	16	O8REQ8 fusobacteri
5	6	37.5	319	5	O9XYQ1 entamoeba h
6	6	37.5	328	5	O9N2T4 caenorhabdi
7	6	37.5	330	5	O23072 caenorhabdi
8	6	37.5	360	16	O87850 streptomyce
9	6	37.5	399	12	O68409 human cytom
10	6	37.5	430	10	O9SMW7 arabidopsis
11	6	37.5	674	12	O9WI36 cauliflowe
12	6	37.5	680	12	O83169 cauliflowe
13	6	37.5	680	12	O66162 cauliflowe
14	6	37.5	1091	16	O9KAQ7 bacillus ha
15	6	37.5	1285	12	O8V3M7 swinepox vi
16	5	31.2	41	16	O9PIF6 campylobact

17	5	31.2	52	8	Q9ZZU6	Q9ZZU6 littorina s
18	5	31.2	57	16	Q8RH01	Q8RH01 fusobacteri
19	5	31.2	83	2	Q9AIF7	Q9AIF7 carsonella
20	5	31.2	90	17	O8TMA6	O8TMA6 methanosarc
21	5	31.2	103	2	Q9Z5P6	Q9Z5P6 enterococcu
22	5	31.2	106	16	O55541	O55541 synechocyst
23	5	31.2	107	10	Q9S6Y4	Q9S6Y4 lotus corni
24	5	31.2	107	10	Q9S6Y3	Q9S6Y3 lotus corni
25	5	31.2	108	2	Q936F3	Q936F3 staphylococ
26	5	31.2	113	2	Q8VVE1	Q8VVE1 uncultured
27	5	31.2	118	2	Q9FDG9	Q9FDG9 streptococc
28	5	31.2	120	16	Q8Y6N6	Q8Y6N6 listeria mo
29	5	31.2	123	2	O68605	O68605 campylobact
30	5	31.2	126	4	O95545	O95545 homo sapien
31	5	31.2	128	2	Q53372	Q53372 synechocyst
32	5	31.2	128	17	O26942	O26942 methanobact
33	5	31.2	134	17	O9HK39	O9HK39 thermoplas
34	5	31.2	136	4	Q13964	Q13964 homo sapien
35	5	31.2	136	5	Q21953	Q21953 caenorhabdi
36	5	31.2	138	16	Q97KA8	Q97KA8 clostridium
37	5	31.2	147	17	Q974Q5	Q974Q5 sulfolobus
38	5	31.2	148	4	Q9NWH4	Q9NWH4 homo sapien
39	5	31.2	155	16	O8REG9	O8REG9 fusobacteri
40	5	31.2	158	2	Q98FA1	Q98FA1 salmonella
41	5	31.2	160	16	Q97L62	Q97L62 clostridium
42	5	31.2	161	16	Q8AKN7	Q8AKN7 clostridium
43	5	31.2	163	16	Q8YPA6	Q8YPA6 anabaena sp
44	5	31.2	173	16	Q99UW3	Q99UW3 staphylococ
45	5	31.2	178	4	O8WXJ2	O8WXJ2 homo sapien
46	5	31.2	180	16	O8U6F5	O8U6F5 agrobacteri
47	5	31.2	181	16	O9POL6	O9POL6 ureaplasma
48	5	31.2	187	16	O9PIA4	O9PIA4 campylobact
49	5	31.2	197	5	Q9GPR4	Q9GPR4 dictyosteli
50	5	31.2	208	16	Q927E8	Q927E8 listeria in
51	5	31.2	208	16	Q8Y3Y6	Q8Y3Y6 listeria mo
52	5	31.2	211	4	Q9NRL8	Q9NRL8 homo sapien
53	5	31.2	212	5	O95VB9	O95VB9 tetrahymena
54	5	31.2	214	5	O20964	O20964 caenorhabdi
55	5	31.2	214	5	O20968	O20968 caenorhabdi
56	5	31.2	214	14	P97966	P97966 unidentified
57	5	31.2	214	16	Q8RD62	Q8RD62 thermosuaer
58	5	31.2	215	11	P97470	P97470 mus musculus
59	5	31.2	216	6	O9N224	O9N224 gorilla gor
60	5	31.2	216	6	O9N222	O9N222 gorilla gor
61	5	31.2	216	6	O9N1Z7	O9N1Z7 pongo pygma
62	5	31.2	217	2	O8RR39	O8RR39 streptococc
63	5	31.2	217	10	Q8SIQ1	Q8SIQ1 oryza sativ
64	5	31.2	219	16	Q9JQX9	Q9JQX9 neisseria m
65	5	31.2	220	16	O8XMG7	O8XMG7 clostridium

## ALIGNMENTS

RESULT 1	Q36097	PRELIMINARY;	PRT;	483 AA.
ID	Q36097			
AC	Q36097			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
DE	Cytochrome c oxidase polypeptide I (BC 1.9.3.1) (Fragment).			
GN	COI.			
OS	Theileria parva.			
OG	Mitochondrion.			
OC	Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileridae;			
OC	Theileria.			
OX	NCBI_TaxID=5875;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MUGUGA;			
RX	MEDLINE=94155854; PubMed=8112303;			
RA	Kairo A., Fairlamb A., Gobright E., Nene V.;			

RT "A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA  
sequences and open reading frames for mitochondrially-encoded  
proteins.";

RL EMBO J. 13:898-905(1994).  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
AND COPPER B (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 4 FERRICYTOCHROME  
C + 2 H(2)O.  
CC -1- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).  
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
INNER MEMBRANE (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
DR EMBL; 223263; CAA80798.1; -  
DR InterPro; IPR000883; COX1.  
DR Pfam; PF00115; COX1; 1.  
KW Copper; Heme; inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; transmembrane.  
FT NON\_TER 1  
SQ SEQUENCE 483 AA; 54008 MW; 314438D6EF4CF3D6 CRC64;

Query Match 43.8%; Score 7; DB 8; Length 483;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFIIIF 12

DB 150 GIDFIIIF 156

RESULT 2

ID O27080 PRELIMINARY; PRT; 156 AA.

AC O27080;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE N-terminal acetyltransferase complex, subunit ARD1.  
GN MTH999  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DELTA H;  
RX MEDLINE=98037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
deltaH: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).

DR EMBL; AE000872; AAB85496.1; -

DR InterPro; IPR000182; GCN5acetyltransf.

DR Pfam; PF00583; Acetyltransf; 1.

KW Transferrase; Complete proteome.

SQ SEQUENCE 156 AA; 18162 MW; 910568F70B30A041 CRC64;

Query Match 37.5%; Score 6; DB 17; Length 156;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIFI 14

DB 53 FIIIFI 58

RESULT 3

O25493 PRELIMINARY; PRT; 205 AA.

ID O25493;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Hypothetical protein HP0813.

GN HP0813

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OX NCBI\_TaxID=210;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-26695 / ATCC 700392;

RA MEDLINE=97394467; PubMed=9252185;

RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,

RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,

RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,

RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,

RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,

RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,

RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,

RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,

RT Venter J.C.;

RT "The complete genome sequence of the gastric pathogen Helicobacter

pylori.";

RL Nature 388:539-547(1997).

DR EMBL; AE000593; AAD07862.1; -

DR TIGR; HP0813; -

DR InterPro; IPR001279; Blactmase-like.

DR Pfam; PF00753; lactamase\_B; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 205 AA; 23463 MW; 9FD355E0B19C2DBE CRC64;

Query Match 37.5%; Score 6; DB 16; Length 205;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11

DB 23 GIDFII 28

RESULT 4

O8RE08 PRELIMINARY; PRT; 250 AA.

ID O8RE08

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Branched-chain amino acid transport protein azlc.

GN FN1039

OS Fusobacterium nucleatum (subsp. nucleatum).

OC Bacteria; Fusobacteria; Fusobacterium.

OX NCBI\_TaxID=76856;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 25586;

RA MEDLINE=21886394; PubMed=11889109;

RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,

RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,

RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,

RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,

RA Fonstein M., Kyrpides N., Overbeek R.;

RT "Genome sequence and analysis of the oral bacterium Fusobacterium

nucleatum strain ATCC 25586.";

RL J. Bacteriol. 184:2005-2018(2002).



DR EMBL; AEO10611; AAL95235.1; -  
KW Complete proteome.  
SQ SEQUENCE 250 AA; 28752 MW; 42F8DF85F7432AF CRC64;  
Query Match 37.5%; Score 6; DB 16; Length 250;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFII 11  
Db 179 GIDFII 184  
RESULT 5  
Q9XYQ1 PRELIMINARY; PRT; 319 AA.  
AC Q9XYQ1  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Inositol 1,3,4-trisphosphate 5/6-kinase.  
GN IK.  
OS Entamoeba histolytica.  
OC Eukaryota; Entamoebidae; Entamoeba.  
OX NCBI\_TaxID=5759;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HM-1:IMSS;  
RX MEDLINE=20264031; PubMed=10802324;  
RA Field J., Wilson M.P., Mai Z., Majerus P.W., Samuelson J.;  
RT "An Entamoeba histolytica inositol 1,3,4-trisphosphate 5/6-kinase has  
a novel 3-kinase activity."  
RL Mol. Biochem. Parasitol. 108:119-123(2000).  
DR EMBL; AFI18848; AAD2969.1; -  
KW Kinase.  
SQ SEQUENCE 319 AA; 36480 MW; D526DBF2E897305D CRC64;  
Query Match 37.5%; Score 6; DB 5; Length 319;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5 CGIDFI 10  
Db 272 CGIDFI 277  
RESULT 6  
Q9N2T4 PRELIMINARY; PRT; 328 AA.  
AC Q9N2T4  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Y9C9A.9 protein.  
GN Y9C9A.9.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC024882; AAF60933.1; -

DR InterPro; IPR003002; 7TM\_chenol.  
DR InterPro; IPR000168; 7TM\_nematode.  
DR Pfam; PF01461; 7tm\_4; 1.  
SQ SEQUENCE 328 AA; 37879 MW; 86DD68EA42EB8FED CRC64;  
Query Match 37.5%; Score 6; DB 5; Length 328;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 IIFWIF 15  
Db 134 IIFWIF 139  
RESULT 7  
Q23072 PRELIMINARY; PRT; 330 AA.  
AC Q23072  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ZC142.1 protein.  
GN ZC142.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,  
Parsons J., Percy C., Rifkin L., Ropra A., Saunders D., Showkeen R.,  
Smaldon N., Smith A., Sonhammer E., Staden R., Suiston J.,  
Therly-Mieg J., Thomas K., Vaughan M., Vaughan K., Waterston R.,  
Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans."  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Bradshaw H.;  
RT "The sequence of C. elegans cosmid ZC142.";  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U64841; AAB04845.1; -  
DR InterPro; IPR002651; DUF32.  
DR Pfam; PF01748; DUF32; 1.  
SQ SEQUENCE 330 AA; 37731 MW; 7C7399A117F7C463 CRC64;  
Query Match 37.5%; Score 6; DB 5; Length 330;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 IIFWIF 15  
Db 133 IIFWIF 138  
RESULT 8  
O87850 PRELIMINARY; PRT; 360 AA.  
ID O87850  
AC O87850;

```

DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative secreted protein.
GN SC06595 OR SC8A5.16.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL031013; CAA19788.1; -.
SQ SEQUENCE 360 AA; 38776 MW; BB1ED7F5BBAD8E7A CRC64;

Query Match 37.5%; Score 6; DB 16; Length 360;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FOANCG 6
Db 90 FOANCG 95

RESULT 9
Q8409
ID Q68409 PRELIMINARY; PRT; 399 AA.
AC Q68409;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Orf UL154.
OS Human cytomegalovirus (strain Towne).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOWNE;
RX MEDLINE=96099416; PubMed=8523595;
RA Cha T.A., Tom E., Kemble G.W., Duke G.M., Mocarski E.S., Spaete R.R.;
RA "Human cytomegalovirus clinical isolates carry at least 19 genes not
RA found in laboratory strains.";
RL J. Virol. 70:78-83(1996).
DR EMBL: U33332; AAA85896.1; -.
SQ SEQUENCE 399 AA; 45181 MW; 6D89F5267EF17998 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 399;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIFI 14
Db 20 FIIIFI 25

RESULT 10
Q9SMW7
ID Q9SMW7 PRELIMINARY; PRT; 430 AA.
AC Q9SMW7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)

```

```

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Hypothetical 48.0 kDa protein.
GN TP19.230.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Choise N., Robert C., Brottier P., Wincker P., Cattolico L.,
RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Quetier F., Salanoubat M.;
RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL133315; CAB62361.1; -.
DR InterPro: IPR003480; Transferase.
DR Pfam: PF02458; Transferase; 1.
KW Hypothetical protein.
SQ SEQUENCE 430 AA; 48004 MW; 6210941B7C148B31 CRC64;

Query Match 37.5%; Score 6; DB 10; Length 430;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ANCGID 8
Db 103 ANCGID 108

RESULT 11
Q9WI36
ID Q9WI36 PRELIMINARY; PRT; 674 AA.
AC Q9WI36;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Reverse transcriptase.
OS Cauliflower mosaic virus.
OC Viruses; Retroviridae; Caulimoviridae; Caulimovirus.
OX NCBI_TaxID=10641;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=XINJIANG;
RA Fang R., Wu X., Bu M., Tian Y., Cai F., Mang K.;
RA "Complete nucleotide sequence of cauliflower mosaic virus (Xinjiang
RA isolate) genomic DNA.";
RL Ping Tu Hsueh Pao 1:247-256(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=XINJIANG;
RA Fang R., Wu X., Bu M., Tian Y., Cai F., Mang K.;
RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF140604; AAD37341.1; -.
DR MEROPS; A03.001; -.
DR InterPro: IPR000588; Peptidase_A3.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF02160; Peptidase_A3; 1.
DR Pfam: PF00078; rvt; 1.
DR PRINTS: PR00731; CAULIMOPTASE.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 674 AA; 78176 MW; 7C9AE48ACDA4D205 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 674;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11
Db 6 GIDFII 11

```

Db 111 GIDFII 116

RESULT 12

Q83169 ID Q83169 PRELIMINARY; PRT; 680 AA.  
AC Q83169;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Reverse transcriptase.  
OS Cauliflower mosaic virus.  
OC Viruses; Retroviral viruses; Caulimoviridae; Caulimovirus.  
ON NCBI\_TaxID=10641;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=941143488; PubMed=8310068;  
RA Chenuault K.D., Melcher U.;  
RT "Cauliflower mosaic virus isolate CMV-1.";  
RL Plant Physiol. 101:1395-1396(1993).  
DR MEROPS; A03.001;  
DR InterPro: IPR000588; Peptidase\_A3.  
DR InterPro: IPR000477; RVase.  
DR Pfam: PF02160; Peptidase\_A3; 1.  
DR Pfam: PF00078; rvt; 1.  
DR PRINTS: PR00731; CAULIMOPTASE.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 680 AA; 78727 MW; ACA8A4B9E8316708 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 680;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11

Db 110 GIDFII 115

RESULT 13

Q66162 ID Q66162 PRELIMINARY; PRT; 680 AA.  
AC Q66162;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE ORF V.  
OS Cauliflower mosaic virus.  
OC Viruses; Retroviral viruses; Caulimoviridae; Caulimovirus.  
ON NCBI\_TaxID=10641;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=B29;  
RA Pique M., Mougout J.L., Geldreich A., Guidauci T., Mesnard J.M.,  
RA Lebeurier G., Yot P.;  
RT "Sequence of a cauliflower mosaic virus strain infecting solanaceous plants.";  
RL Gene 155:305-306(1995).  
DR EMBL; X79465; CAA55974.1;  
DR MEROPS; A03.001;  
DR InterPro: IPR000588; Peptidase\_A3.  
DR InterPro: IPR000477; RVase.  
DR Pfam: PF02160; Peptidase\_A3; 1.  
DR Pfam: PF00078; rvt; 1.  
DR PRINTS: PR00731; CAULIMOPTASE.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 680 AA; 78728 MW; CF1EDBBE45EC2A6 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 680;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11

Db 110 GIDFII 115

RESULT 14

Q9KA07 ID Q9KA07 PRELIMINARY; PRT; 1091 AA.  
AC Q9KA07;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein BH2230.  
GN BH2230.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirana C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001514; BAB05949.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1091 AA; 127442 MW; FD95071E13701331 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11

Db 880 GIDFII 885

RESULT 15

Q8V3M7 ID Q8V3M7 PRELIMINARY; PRT; 1285 AA.  
AC Q8V3M7;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE SPV068 RNA polymerase subunit RPO147.  
GN SPV068.  
OS Swinepox virus.  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Suipoxvirus.  
OX NCBI\_TaxID=10276;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=17077-99;  
RC MEDLINE=21624277; PubMed=11752168;  
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Osorio F.A., Balinsky C.,  
RA Kutish G.F., Rock D.L.;  
RT "The genome of swinepox virus.";  
RL J. Virol. 76:783-790(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17077-99;  
RA Afonso C.L., Tulman E.R., Lu Z., Balinsky C., Osorio F.A., Zsak L.,  
RA Kutish G.F., Rock D.L.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF410153; AAL69807.1;  
DR InterPro: IPR000722; RNA\_pol\_A.  
DR Pfam; PF00623; RNA\_pol\_A; 1.  
SQ SEQUENCE 1285 AA; 147680 MW; 1D34EFF8CD873DC6 CRC64;

Query Match 37.5%; Score 6; DB 12; Length 1285;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 11  
 Db 552 GIDFII 557

## RESULT 16

Q9PIF6 PRELIMINARY; PRT; 41 AA.  
 AC Q9PIF6;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical protein Cj0344.  
 GN Cj0344.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NTCT 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrrell B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 DR EMBL; AL139075; CAB74181.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 41 AA; 5107 MW; 2698D105B0F0BF04 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12  
 Db 10 DFIIIF 14

## RESULT 17

Q9Z2U6 PRELIMINARY; PRT; 52 AA.  
 AC Q9Z2U6;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE ATPase subunit 8.  
 GN ATPase 8.  
 OS Littorina saxatilis.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;  
 OC Mesogastropoda; Littorinoidea; Littorinidae; Littorina.  
 OX NCBI\_TaxID=31220;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9919185; PubMed=10093225;  
 RA Wilding C.S., Mill P.J., Graham J.;  
 RT "Partial sequence of the mitochondrial genome of Littorina saxatilis:  
 RT relevance to gastropod phylogenetics.";  
 RL J. Mol. Evol. 48:348-359(1999).  
 RL EMBL; AJ132137; CAA10595.1; -;  
 DR InterPro; IPR001421; ATPase8\_mit.  
 DR Pfam; PF00895; ATP-synt\_8; 1.  
 KW Mitochondrion.  
 SQ SEQUENCE 52 AA; 6237 MW; 756DDE254F629499 CRC64;

Query Match 31.2%; Score 5; DB 8; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIIFW 13  
 Db 13 FIIIFW 17

## RESULT 18

Q8RH01 PRELIMINARY; PRT; 57 AA.  
 AC Q8RH01;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein FN0120.  
 GN FN0120.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteria; Fusobacterium.  
 OX NCBI\_TaxID=76856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 25586;  
 RX MEDLINE=21886394; PubMed=11889109;  
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
 RA Fongstein M., Kyrpides N., Overbeek R.;  
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 RT nucleatum strain ATCC 25586.";  
 RL J. Bacteriol. 184:2005-2018(2002).  
 DR EMBL; AE010526; AAL94329.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 57 AA; 6631 MW; EF675C73A500BE CRC64;

Query Match 31.2%; Score 5; DB 16; Length 57;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFII 10  
 Db 21 GIDFII 25

## RESULT 19

Q9AIF7 PRELIMINARY; PRT; 83 AA.  
 AC Q9AIF7;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Ribosomal protein S17.  
 GN RPS17.  
 OS Carsonella ruddii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.  
 OX NCBI\_TaxID=114186;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21125546; PubMed=11222582;  
 RA Clark M.A., Baumann L., Thao M.L., Moran N.A., Baumann P.;  
 RT "Degenerative Minimalism in the Genome of a Psyllid Endosymbiont.";  
 RL J. Bacteriol. 183:1853-1861(2001).  
 DR EMBL; AF274444; AAK17089.1; -;  
 DR HSP; P23828; IRIP.  
 DR InterPro; IPR000266; Ribosomal\_S17.  
 DR Pfam; PF00366; Ribosomal\_S17; 1.  
 DR ProDom; PD001295; Ribosomal\_S17; 1.  
 SQ SEQUENCE 83 AA; 10205 MW; 72EA2AC8F1AC09A8 CRC64;

Query Match 31.2%; Score 5; DB 2; Length 83;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14  
Db 68 IIFWI 72

## RESULT 20

Q8TMA6 PRELIMINARY; PRT; 90 AA.  
AC Q8TMA6;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Predicted protein.  
GN MA2759.  
OS Methanosarcina acetivorans.  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2214;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
RX MEDLINE=21929760; PubMed=11932238;  
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,  
RA Allen N., Naylor J., Stange-Thomann N., DeAtrellano K., Johnson R.,  
RA Linton L., McWarren P., McKernan K., Talamas J., Tirrell A., Ye W.,  
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,  
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,  
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.J.L., Paulsen I.,  
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
RA Metcalf W.W., Birren B.;  
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
and physiological diversity.";  
RL Genome Res. 12:532-542(2002).  
DR EMBL; AE010975; AAM06137.1; -;  
KW Complete proteome.  
SQ SEQUENCE 90 AA; 10457 MW; DDD0D4BC3F000CD8 CRC64;

Query Match 31.2%; Score 5; DB 17; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 IIFWI 13  
Db 75 IIFWI 79

## RESULT 21

Q9Z5P6 PRELIMINARY; PRT; 103 AA.  
AC Q9Z5P6;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE Hypothetical 12.0 kDa protein (Trf protein).  
GN EP0045.  
OS Enterococcus faecalis (Streptococcus faecalis).  
OC Plasmid pAM373.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Enterococcaceae; Enterococcus.  
OX NCBI\_TaxID=1351;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC PLASMID=pAM373;  
RA Muscholl-Silberhorn A.B.;  
RT "Cloning and expression of Asa373, a novel adhesin unrelated to the  
other sex pheromone plasmid-encoded aggregation substances of  
Enterococcus faecalis.";

Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE FROM N.A.

RC PLASMID=pAM373;  
RX MEDLINE=20453452; PubMed=10998166;  
RA De Boever E.H., Clewell D.B., Fraser C.M.;  
RT "Enterococcus faecalis conjugative plasmid pAM373: complete nucleotide  
sequence and genetic analyses of sex pheromone response.";  
RL Mol. Microbiol. 37:1327-1341(2000).  
DR EMBL; AJ132039; CAB38224.1; -;  
DR EMBL; AE002565; AAG40456.1; -;  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 103 AA; 12019 MW; 91F4F000B80B2D35 CRC64;

Query Match 31.2%; Score 5; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCGID 8  
Db 26 NCGID 30

## RESULT 22

Q55541 PRELIMINARY; PRT; 106 AA.  
AC Q55541;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein slr0333.  
GN SLR0333.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PCC 6803;  
RA Tabata S.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
RA Sugiyama M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
Synechocystis sp. strain PCC6803. II. Sequence determination of the  
entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL; D63999; BAA10038.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 106 AA; 11406 MW; EEE4CF44B5B79C3B CRC64;

Query Match 31.2%; Score 5; DB 16; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
Db 97 DFIIF 101

## RESULT 23

Q9S6Y4 PRELIMINARY; PRT; 107 AA.  
 AC Q9S6Y4;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Dihydroflavanol reductase 3 (Fragment).  
 GN DFR.  
 OS Lotus corniculatus (Bird's-foot trefoil).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.  
 OX NCBI\_TaxID=47247;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LEAF;  
 RA Paolucci F., Capucci R., Arcioni S., Damiani F.;  
 RT "Birdsfoot trefoil: a model for studying the synthesis of condensed  
 RT tannin.";  
 RL (In) Gross G.G., Hemingway R.W., Yoshida T. (eds.);  
 RL PLANT POLYPHENOLS 2, CHEMISTRY AND BIOLOGY, pp.1-1, Plenum Press,  
 RL New York, NY, USA (2000).  
 DR EMBL; AF117263; AAF23884.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 107  
 SQ SEQUENCE 107 AA; 12242 MW; C5BC292A44246D7D CRC64;

Query Match 31.2%; Score 5; DB 10; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 Db 86 GIDFI 90

## RESULT 24

Q9S6Y3 PRELIMINARY; PRT; 107 AA.  
 AC Q9S6Y3;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Dihydroflavanol reductase 4 (Fragment).  
 GN DFR.  
 OS Lotus corniculatus (Bird's-foot trefoil).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.  
 OX NCBI\_TaxID=47247;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LEAF;  
 RA Paolucci F., Capucci R., Arcioni S., Damiani F.;  
 RT "Birdsfoot trefoil: a model for studying the synthesis of condensed  
 RT tannin.";  
 RL (In) Gross G.G., Hemingway R.W., Yoshida T. (eds.);  
 RL PLANT POLYPHENOLS 2, CHEMISTRY AND BIOLOGY, pp.1-1, Plenum Press,  
 RL New York, NY, USA (2000).  
 DR EMBL; AF117264; AAF23885.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 107  
 SQ SEQUENCE 107 AA; 12170 MW; C6E1C4A44246D7D CRC64;

Query Match 31.2%; Score 5; DB 10; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 Db 86 GIDFI 90

## RESULT 25

Q936F3 PRELIMINARY; PRT; 108 AA.  
 ID Q936F3;  
 AC Q936F3;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Hypothetical 12.2 kDa protein.  
 GN DE.  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 OX NCBI\_TaxID=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M;  
 RA Luong T.T., Shu O., Bush K., Lee C.Y.;  
 RT "The Type1 Capsular Polysaccharide of Staphylococcus aureus is carried  
 RT in a Staphylococcal Cassette Chromosome Genetic Element.";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U10927; AAL26685.1; -.  
 GN Hypothetical protein.  
 KW SEQUENCE 108 AA; 12234 MW; 40380A7E555A40DB CRC64;

Query Match 31.2%; Score 5; DB 2; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 Db 34 GIDFI 38

## RESULT 26

Q8VVE1 PRELIMINARY; PRT; 113 AA.  
 ID Q8VVE1;  
 AC Q8VVE1;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Ssb protein.  
 GN SSB.  
 OS uncultured bacterium.  
 OG Plasmid pEMT3.  
 OC Bacteria; environmental samples.  
 OX NCBI\_TaxID=77133;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gstaider M.E.;  
 RT "Caractérisation de plasmides a large spectre d'hotels isoles de  
 RT biotopes pollués.";  
 RL Thesis (2001), Department of Biological sciences,  
 RL Université Libre de Bruxelles, Bruxelles, Belgium.  
 DR EMBL; AJ414161; CAC94917.1; -.  
 DR InterPro; IPR000424; SSB\_protein.  
 DR Pfam; PF00436; SSB; 1.  
 DR TIGRFAMs; TIGR00621; ssb; 1.  
 KW Plasmid.  
 SQ SEQUENCE 113 AA; 12836 MW; 9F5D8B9ABF476FE CRC64;

Query Match 31.2%; Score 5; DB 2; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GIDFI 10  
 |||||  
 Db 94 GIDFI 98

## RESULT 27

Q9FDG9 PRELIMINARY; PRT; 118 AA.  
 ID Q9FDG9

Q9FDG9;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical 13.5 kDa protein.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CH43;  
 RX MEDLINE=20378620; PubMed=10919773;  
 RA Qi F., Chen P., Caulfield P.W.;  
 RT "Purification and biochemical characterization of mutacin I from the  
 RT group I strain of streptococcus mutans, CH43, and genetic analysis of  
 RT mutacin I biosynthesis genes.";  
 RL Appl. Environ. Microbiol. 66:3221-3229(2000).  
 DR EMBL; AF267498; AAF99696.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 118 AA; 13465 MW; CID9590A77CF4D4D CRC64;  
 Query Match 31.2%; Score 5; DB 2; Length 118;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 11 IFWIF 15  
 Db 14 IFWIF 18  
 RESULT 28  
 Q8Y6N6 PRELIMINARY; PRT; 120 AA.  
 AC Q8Y6N6;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein Iml648.  
 GN LMO1648.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EGD-E / SEROVAR 1/2A;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurupkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 RT "Comparative genomics of Listeria species.";  
 RL Science 294:849-852(2001).  
 DR EMBL; AL591980; CAC99726.1; -;  
 DR ListList; LMO01648; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 120 AA; 14245 MW; AE139D9B67330B8F CRC64;  
 Query Match 31.2%; Score 5; DB 16; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 IDFII 11  
 Db 27 IDFII 31

RESULT 29  
 O68605 PRELIMINARY; PRT; 123 AA.  
 AC O68605;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 13.5 kDa protein (Fragment).  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC11168;  
 RX MEDLINE=98440418; PubMed=9765558;  
 RA van Vliet A.H.M., Wooldridge K.G., Ketley J.M.;  
 RT "Iron-responsive gene regulation in a Campylobacter jejuni fur  
 RT mutant.";  
 RL J. Bacteriol. 180:5291-5298(1998).  
 DR EMBL; AF052056; AAC64261.1; -;  
 DR InterPro; IPR003825; Colicin\_V.  
 DR Pfam; PF02674; Colicin\_V; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 123 AA; 13508 MW; 2B2F9B250E4B63A7 CRC64;  
 Query Match 31.2%; Score 5; DB 2; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 IIFWI 14  
 Db 13 IIFWI 17  
 RESULT 30  
 O95545 PRELIMINARY; PRT; 126 AA.  
 AC O95545;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE DJ283E3.6.3 (Putative novel protein similar to many (Archae)bacterial,  
 DE worm and yeast hypothetical proteins) (Putative partial isoform 3)  
 DE (Fragment).  
 GN DJ283E3.6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pearce A.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL031282; CAA20355.1; -;  
 DR InterPro; IPR002504; ATP\_NADK.  
 DR Pfam; PF01513; NAD\_Kinase; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 126 AA; 14188 MW; 79D000A1E5B499B1 CRC64;  
 Query Match 31.2%; Score 5; DB 4; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 IDFII 11  
 Db 96 IDFII 100

RESULT 31  
Q53372  
ID Q53372 PRELIMINARY; PRT; 128 AA.  
AC Q53372;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE C-type cytochrome.  
GN CYTM  
OS Synechocystis sp. (strain PCC 6714).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1147;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PCC6714;  
RA Malakhov M.P., Semenenko V.E.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
DR HSSP; X82563; CAAS57910.1;  
DR HSSP; P00044; ICRI.  
DR InterPro; IPR000345; CytC\_heme\_bind.  
DR Pfam; PF00034; Cytochrome\_C; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
FT CHAIN 53 128  
FT POTENTIAL.  
SQ SEQUENCE 128 AA; 13848 MW; D7725CD852C25C01 CRC64;

Query Match 31.2%; Score 5; DB 2; Length 128;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FQANC 5  
Db 64 FQANC 68

RESULT 32  
Q26942  
ID Q26942 PRELIMINARY; PRT; 128 AA.  
AC Q26942;  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE FERREDOXIN.  
GN MTH854.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DELTA H;  
RA MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Frabnakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
DeltaH: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL; AE000862; AB88352.1;  
DR HSSP; P00198; 2FDN.  
DR InterPro; IPR001450; 4Fe4S\_Ferredoxin.  
DR Pfam; PF00037; fer4; 2.  
DR PRINTS; PR00353; 4FE4SFERDOXIN.  
DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 2.  
KW Iron-sulfur; Complete proteome.  
SQ SEQUENCE 128 AA; 14216 MW; 4EEB2C4261D5BE9F CRC64;

Query Match 31.2%; Score 5; DB 17; Length 128;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GIDFI 10  
Db 54 GIDFI 58  
|||||

## RESULT 33

Q9HK39  
ID Q9HK39 PRELIMINARY; PRT; 134 AA.  
AC Q9HK39;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)  
DE Hypothetical protein Ta0768.  
GN TA0768.  
OS Thermoplasma acidophilum.  
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
OC Thermoplasmataceae; Thermoplasma.  
OX NCBI\_TaxID=2303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 1728;  
RA MEDLINE=20479972; PubMed=11029001;  
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
acidophilum.";  
RL Nature 407:508-513(2000).  
DR EMBL; AL445065; CAC11900.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 134 AA; 15217 MW; 5A3A02E8FB8FF4FE CRC64;

Query Match 31.2%; Score 5; DB 17; Length 134;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 127 FIIFW 131  
|||||

## RESULT 34

Q13964  
ID Q13964 PRELIMINARY; PRT; 136 AA.  
AC Q13964;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE Glycoprotein CD28.  
GN CD28.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9608;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=90293482; PubMed=2162892;  
RA Lee K.P., Taylor C., Petryniak B., Turka L.A., June C.H.,  
RA Thompson C.B.;  
RT "The genomic organization of the CD28 gene. Implications for the  
regulation of CD28 mRNA expression and heterogeneity.";  
RL J. Immunol. 145:344-352(1990).  
DR EMBL; M37815; AAA51945.1;  
DR EMBL; M37812; AAA51945.1; JOINED.  
DR EMBL; M37813; AAA51945.1; JOINED.  
DR EMBL; M37814; AAA51945.1; JOINED.  
SQ SEQUENCE 136 AA; 15369 MW; C9AF33467706D2BE CRC64;

Query Match 31.2%; Score 5; DB 4; Length 136;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```
QY 9 FIIFW 13
Db 90 FIIFW 94

RESULT 35
Q21953
ID Q21953 PRELIMINARY; PRT; 136 AA.
AC Q21953;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 15.2 kDa protein.
GN R12B2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditiodea;
OC Rhabditiidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Miller N.;
RT "The sequence of C. elegans cosmid R12B2."
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U00066; AAA50741.1; -
KW Hypothetical protein.
SQ SEQUENCE 136 AA; 15226 MW; 611D037F55CA7696 CRC64;

Query Match 31.2%; Score 5; DB 5; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12
Db 25 DFIIIF 29

RESULT 36
Q97KA8
ID Q97KA8 PRELIMINARY; PRT; 138 AA.
AC Q97KA8;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Uncharacterized small membrane protein, homolog of ykva
DE B.subtilis.
GN CAC1011.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;

RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007616; AAK78987.1; -
KW Complete proteome.
SQ SEQUENCE 138 AA; 16257 MW; 570837AF9060ACC5 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWI 14
Db 117 IIFWI 121

RESULT 37
Q974Q5
ID Q974Q5 PRELIMINARY; PRT; 147 AA.
AC Q974Q5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ST0604.
GN ST0604.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
Crenarchaeon, Sulfolobus tokodaii strain 7."
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000983; BAB65602.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 147 AA; 17362 MW; DE2A7D819E08A3C6 CRC64;

Query Match 31.2%; Score 5; DB 17; Length 147;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIIF 12
Db 36 DFIIIF 40

RESULT 38
Q9NWH4
ID Q9NWH4 PRELIMINARY; PRT; 148 AA.
AC Q9NWH4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 16.9 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EMBRYO;
RA Isogai T., Oka T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
```

RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK000886; BAA91406.1;  
 DR InterPro; IPR001081; Ribosomal\_L20.  
 DR ProbDom; PD002389; Ribosomal\_L20; 1.  
 SQ SEQUENCE 148 AA; 16903 MW; 9A28D55AEDEB4766 CRC64;

Query Match 31.2%; Score 5; DB 4; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QANCG 6  
 Db 117 QANCG 121

RESULT 39  
 Q8REG9 PRELIMINARY; PRT; 155 AA.  
 ID Q8REG9  
 AC Q8REG9;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical cytosolic protein FN1134.  
 GN FN1134.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteria; Fusobacterium.  
 OX NCBI\_TaxID=76856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 25586;  
 RX MEDLINE=21886394; PubMed=11889109;  
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
 RA Feinstein M., Kyriplides N., Overbeek R.,  
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 RT nucleatum strain ATCC 25586."  
 RL J. Bacteriol. 184:2005-2018(2002).  
 RL EMBL; AE010618; AAL95330.1;  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 155 AA; 18101 MW; CB3EDB741E897B97 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 155;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 IDFTI 11  
 Db 27 IDFTI 31

RESULT 40  
 Q9RFAL PRELIMINARY; PRT; 158 AA.  
 ID Q9RFAL  
 AC Q9RFAL;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 17.3 kDa protein (Fragment).  
 OS Salmonella pullorum.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OX Salmonella  
 OX NCBI\_TaxID=605;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Patton T., Tseng C.W., Sobieski R.J., Crupper S.S.;  
 RT "Salmonella pullorum chromosomal region demonstrating homology to  
 RT Escherichia coli K-12 MG1655 section 398 to 400 of the genome."  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF198630; AAF15288.1;  
 DR InterPro; IPR002599; Metalloenzyme.  
 DR Pfam; PF01676; Metalloenzyme; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 FT NON\_TER 158  
 SQ SEQUENCE 158 AA; 17266 MW; 96F489ADBDFEB862 CRC64;

Query Match 31.2%; Score 5; DB 2; Length 158;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ANCGI 7  
 Db 31 ANCGI 35

RESULT 41  
 Q97L62 PRELIMINARY; PRT; 160 AA.  
 ID Q97L62  
 AC Q97L62;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Predicted tRNA-methylase (SpoU class).  
 GN CAC0700.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 OX NCBI\_TaxID=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=21359325; PubMed=11466286;  
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium Clostridium acetobutylicum."  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AE007585; AAK78677.1;  
 DR InterPro; IPR001537; SpoU\_methylase.  
 DR Pfam; PF00588; SpoU\_methylase; 1.  
 KW Methyltransferase; Complete proteome.  
 SQ SEQUENCE 160 AA; 18424 MW; 2F6812B632847820 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 DFIF 12  
 Db 101 DFIF 105

RESULT 42  
 Q8XKN7 PRELIMINARY; PRT; 161 AA.  
 ID Q8XKN7  
 AC Q8XKN7;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Probable transcriptional regulator.  
 GN CPE1357.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 OX NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-13 / TYPE A;  
 RX PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,  
RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
RT flesh-eater.",  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
DR EMBL: AP003190; BAB81063.1; -.  
KW Complete proteome.  
SQ SEQUENCE 161 AA; 17955 MW; 23772A02B24563B2 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 161;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DFIIF 12  
Db 128 DFIIF 132

RESULT 43  
Q8YPA6 PRELIMINARY; PRT; 163 AA.  
AC Q8YPA6;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein Alr4293.  
GN Alr4293.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriuchimi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpou S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.",  
RL DNA Res. 8:205-213(2001).  
DR EMBL: AP003596; BAB75992.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 163 AA; 18352 MW; AF0D6B94CACD6806 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 IFWIF 15  
Db 58 IFWIF 62

RESULT 44  
Q99UW3 PRELIMINARY; PRT; 173 AA.  
AC Q99UW3;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein SAV1142.  
GN SAV1142 OR SA0989.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
OS Staphylococcus aureus (strain N315).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Staphylococcus.  
OX NCBI\_TaxID=158878; 158879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mitutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaiko C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus.",  
RL Lancet 357:1225-1240(2001).  
DR EMBL: AP003361; BAB57304.1; -.  
DR EMBL: AP003132; BAB42238.1; -.  
DR InterPro: IPR003825; Colicin\_V.  
DR Pfam: PF02674; Colicin\_V.1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 173 AA; 20286 MW; E8AD9D5C728EB44E CRC64;

Query Match 31.2%; Score 5; DB 16; Length 173;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IDFIIF 11  
Db 3 IDFIIF 7

RESULT 45  
Q8WXJ2 PRELIMINARY; PRT; 178 AA.  
AC Q8WXJ2;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE CD28 antigen (Fragment).  
GN CD28.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21601151; PubMed=11735222;  
RA Ling V., Wu P.W., Finnerty H.F., Agostino M.J., Graham J.R., Chen S.,  
RA Jussiff J., Fisk G.J., Miller C.P., Collins M.;  
RT "Assembly and Annotation of Human Chromosome 2q33 Sequence Containing  
RT the CD28, CTLA4, and ICOS Gene Cluster: Analysis by Computational,  
RT Comparative, and Microarray Approaches.",  
RL Genomics 78:155-168(2001).  
DR EMBL: AF411057; AAL40931.1; -.  
DR InterPro: IPR003600; Ig\_Like.  
DR SMART: SM00410; Ig\_Like; 1.  
FT NON\_TER 178 178  
SQ SEQUENCE 178 AA; 20022 MW; D60EBA4174AAABD0 CRC64;

Query Match 31.2%; Score 5; DB 4; Length 178;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FIIFW 13  
Db 174 FIIFW 178

Search completed: May 11, 2003, 20:12:12  
Job time : 11.4956 secs



GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 13:47:50 ; Search time 362.761 Seconds  
(without alignments)  
1283.613 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FOANCCIDFIIFWIFW 16

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Egapop 6.0 , Egapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_173955\_334/app\_query.fasta\_1.462  
-DB=GenEmbl -OFMT=fastap -SUFFIX=xge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133 -CGN\_1\_1758 -runat\_05052003\_173955\_334 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_ov:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pi:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*

- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_higo\_hum:\*
- 40: em\_higo\_mus:\*
- 41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	98	100.0	337	6	AX321911	AX321911 Sequence
2	98	100.0	2239	6	AX321909	AX321909 Sequence
3	98	100.0	5981	6	AX321910	AX321910 Sequence
4	98	100.0	161280	9	AC093903	AC093903 Homo sapi
5	66	67.3	41179	3	AF003739	AF003739 Caenorhab
6	66	67.3	201746	2	AC006895	AC006895 Caenorhab
7	61	62.2	24259	6	AX251448	AX251448 Sequence
8	60	61.2	94703	2	AC008563	AC008563 Homo sapi
9	60	61.2	110879	9	AC016603	AC016603 Homo sapi
10	60	61.2	204843	9	AC010359	AC010359 Homo sapi
11	59	60.2	67679	2	AC113545	AC113545 Mus muscu
12	59	60.2	161334	10	AC091629	AC091629 Mus muscu
13	59	60.2	198281	10	AL593843	AL593843 Mouse DNA
14	59	60.2	218188	2	AL845270	AL845270 Mus muscu
15	57	58.2	149765	9	AC004984	AC004984 Homo sapi
16	57	58.2	177147	9	AC007539	AC007539 Homo sapi
17	57	58.2	191844	2	AC122410	AC122410 Mus muscu
18	57	58.2	234096	2	AC121818	AC121818 Mus muscu
19	56.5	57.7	236925	2	AC102650	AC102650 Mus muscu
20	56	57.1	1669	10	BC009142	BC009142 Mus muscu
21	56	57.1	22211	9	AL359849	AL359849 Human DNA
22	56	57.1	47887	2	AC015109	AC015109 Drosophill
23	56	57.1	52356	2	AC020166	AC020166 Drosophill
24	56	57.1	119633	2	AC119679	AC119679 Homo sapi
25	56	57.1	119394	2	AC010046	AC010046 Drosophill
26	56	57.1	138098	10	AL603664	AL603664 Mouse DNA
27	56	57.1	165330	2	AC016011	AC016011 Homo sapi
28	56	57.1	171376	3	AC007929	AC007929 Drosophill
29	56	57.1	174673	9	AC019187	AC019187 Homo sapi
30	56	57.1	179055	3	AC093551	AC093551 Drosophill
31	56	57.1	187192	2	AC128532	AC128532 Rattus no
32	56	57.1	187756	2	AC108826	AC108826 Mus muscu
33	56	57.1	190153	3	AC008201	AC008201 Drosophill
34	56	57.1	191240	2	AC027738	AC027738 Homo sapi
35	56	57.1	199851	2	AC125080	AC125080 Mus muscu
36	56	57.1	203926	2	AC011122	AC011122 Homo sapi
37	56	57.1	204207	2	AC023147	AC023147 Homo sapi
38	56	57.1	220842	3	AE003745	AE003745 Drosophill
39	56	57.1	232719	2	AC111123	AC111123 Mus muscu
40	56	57.1	322667	3	AE003592	AE003592 Drosophill
41	55.5	56.6	6526	6	AX344899	AX344899 Sequence
42	55.5	56.6	6526	6	AX347139	AX347139 Sequence
43	55.5	56.6	156146	2	AL139180	AL139180 Homo sapi
44	55.5	56.6	180680	9	AL137141	AL137141 Human DNA
45	55.5	56.6	211624	10	AC114817	AC114817 Mus muscu

ALIGNMENTS

```

AX321911
LOCUS AX321911 337 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 442 from Patent WO0172295.
ACCESSION AX321911
VERSION AX321911.1 GI:17906521
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indrias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 442 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source
Location/Qualifiers
1..337
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 103 a 60 c 93 g 81 t
ORIGIN
1
2 88e-08 Length: 337
Pred. No.: 2 88e-08 Length: 337
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX321911 (1-337)
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
|||||
Db 107 TTCCAGGCCAATTGTGGCATAGATTATCATATTCGGATTTTGG 154
|||||

RESULT 2
LOCUS AX321909 2239 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 440 from Patent WO0172295.
ACCESSION AX321909
VERSION AX321909.1 GI:17906515
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indrias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 440 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source
Location/Qualifiers
1..2239
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 619 a 444 c 493 g 683 t
ORIGIN
1
2 5e-07 Length: 2239
Pred. No.: 2 5e-07 Length: 2239
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX321909 (1-2239)
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
|||||
Db 107 TTCCAGGCCAATTGTGGCATAGATTATCATATTCGGATTTTGG 154
|||||

RESULT 3
LOCUS AX321910 5981 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 441 from Patent WO0172295.
ACCESSION AX321910
VERSION AX321910.1 GI:17906518
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R.,
Indrias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M.,
Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0172295-A 441 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source
Location/Qualifiers
1..5981
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1858 a 1029 c 1098 g 1996 t
ORIGIN
1
7 68e-07 Length: 5981
Pred. No.: 7 68e-07 Length: 5981
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-854-133-587 (1-16) x AX321910 (1-5981)
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
|||||
Db 102 TTCCAGGCCAATTGTGGCATAGATTATCATATTCGGATTTTGG 149
|||||

RESULT 4
LOCUS AC093903 161280 bp DNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-733C7 from 4, complete sequence.
ACCESSION AC093903
VERSION AC093903.3 GI:15920156
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 161280)
AUTHORS Radionenko,M. and Kozlowicz,A.
TITLE The sequence of Homo sapiens BAC clone RP11-733C7
JOURNAL Unpublished (2001)
MEDLINE 3 (bases 1 to 161280)
PUBMED 99063792
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
9847074
REFERENCE 2 (bases 1 to 161280)
AUTHORS Radionenko,M. and Kozlowicz,A.
TITLE The sequence of Homo sapiens BAC clone RP11-733C7
JOURNAL Unpublished (2001)
MEDLINE 3 (bases 1 to 161280)
PUBMED 99063792
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
9847074
Direct Submission
Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 161280)
Waterston,R.H.
Direct Submission
Submitted (04-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 161280)
Waterston,R.
Direct Submission
Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 4, 2001 this sequence version replaced gi:15625016.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
-----
Summary Statistics
Center project name: H_NH0733C07
Drafting Center: WIBR
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-310A13. Actual start of
this clone is at base position 1 of RP11-733C7; actual end is at
base position 161280 of RP11-733C7.

Data from AC009792 was used to finish the clone, AC055827.

The sequence of AC055827 has been incorporated into AC093903.
-----
Location/Qualifiers
1. 161280
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-733C7"
/clone_lib="RPCI-11"
148. 304
/rpt_family="L1"
306. 406
/rpt_family="L1"
471. 525
/rpt_family="L1"
/rpt_family="AT_rich"

repeat_region
repeat_region
repeat_region

FEATURES
source
repeat_region
repeat_region
repeat_region
```

misc\_feature /note="similar to EST BF095483 (NID:g10901193)"  
 17772..18329  
 repeat\_region /note="similar to EST BG181259 (NID:g13703031)"  
 17845..17869  
 repeat\_region /rpt\_family="(CTG)n"  
 18792..18828  
 repeat\_region /rpt\_family="AT\_rich"  
 20734..20784  
 repeat\_region /rpt\_family="MERL\_type"  
 20832..20944  
 repeat\_region /rpt\_family="MERL\_type"  
 20948..21964  
 repeat\_region /rpt\_family="L1"  
 21964..21986  
 repeat\_region /rpt\_family="AT\_rich"  
 21965..22276  
 repeat\_region /rpt\_family="Alu"  
 22277..22323  
 repeat\_region /rpt\_family="L1"  
 22341..22681  
 repeat\_region /rpt\_family="L1"  
 22685..22824  
 repeat\_region /rpt\_family="Alu"  
 22805..22824  
 repeat\_region /rpt\_family="(A)n"  
 22826..23575  
 repeat\_region /rpt\_family="L1"  
 23591..23735  
 repeat\_region /rpt\_family="MERL\_type"

## Alignment Scores:

Pred. No.:	3.31e-05	Length:	161280
Score:	98.00	Matches:	16
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-854-133-587 (1-16) x AC093903 (1-161280)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
 DB 27653 TTTCCAGGCCAATGTGGCATAGTTTATCATATCTGGATTTCGG 27606

RESULT 5  
 LOCUS AF003739 41179 bp DNA linear INV 19-APR-2002  
 DEFINITION Caenorhabditis elegans cosmid M01D7, complete sequence.  
 ACCESSION AF003739  
 VERSION AF003739.2 GI:15145732  
 KEYWORDS HTG.  
 SOURCE Caenorhabditis elegans.

ORGANISM  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.  
 1 (bases 1 to 41179)  
 Waterston,R.  
 GenBank  
 The sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium  
 Science 282 (5396), 2012-2018 (1998)  
 99069613  
 PUBMED 9851916  
 REFERENCE 2 (bases 1 to 41179)  
 Gattung,S. and Goela,D.  
 The sequence of C. elegans cosmid M01D7  
 Unpublished (2001)  
 3 (bases 1 to 41179)  
 Waterston,R.  
 Direct Submission  
 Submitted (12-MAY-1997) Genome Sequencing Center, Washington  
 University  
 4 (bases 1 to 41179)  
 Waterston,R.

## TITLE

JOURNAL

## REFERENCE

AUTHORS

TITLE

JOURNAL

## REFERENCE

AUTHORS

TITLE

JOURNAL

## COMMENT

Direct Submission  
 Submitted (09-AUG-2001) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
 Louis, MO 63110, USA  
 5 (bases 1 to 41179)  
 Waterston,R.  
 Direct Submission  
 Submitted (11-OCT-2001) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
 Louis, MO 63110, USA  
 6 (bases 1 to 41179)  
 Waterston,R.  
 Direct Submission  
 Submitted (19-APR-2002) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
 Louis, MO 63110, USA  
 On Aug 9, 2001 this sequence version replaced g1:2105482.  
 Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RQ, England  
 email: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we provide a small overlap between  
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
 regions were double stranded, sequenced with an alternate chemistry  
 or covered by high quality data (i.e., phred quality >= 30); an  
 attempt was made to resolve all sequencing problems, such as  
 compressions and repeats; all regions were covered by sequence from  
 more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its  
 analysis see:  
 (www.wormbase.org/db/seq/sequence?name=M01D7;class=Sequence)

## NEIGHBORING COSMID INFORMATION

The 5' cosmid is C53H9, 3300 bp overlap; the 3' cosmid is Y71G12A,  
 3500 bp overlap. Actual start of this cosmid is at base position  
 3297 of M01D7; actual end is at 41179 of M01D7.

## NOTES:

Coding sequences below are the result of integration and manual  
 review of the following data: computer analysis using the program  
 Genefinder (P. Green and L. Hillier, personal communication), the  
 large scale EST projects of Yuji Kohara  
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE\_INDEX.html) and The C.  
 elegans ORFome cloning project (http://wofdb.dfci.harvard.edu/),  
 similarity to other proteins from BlastX analyses  
 (http://blast.wustl.edu/), sequence conservation with C. briggsae  
 using Jim Kent's WABA alignment program (Genome Research  
 10:1115-1125, 2000), individual C. elegans GenBank submissions,  
 and personal communications with C. elegans researchers. TRNAS  
 are predicted using the program tRNAscan-SE (Lowe, T.M. and  
 Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

## FEATURES

Source

gene

/organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /chromosome="I"  
 /clone="M01D7"  
 52..5004  
 /gene="egl-30"  
 /note="for a graphical representation of this gene see:  
 (www.wormbase.org/db/seq/sequence?name=M01D7.7a;class=Seq



```

/protein_id="AAB58065.1"
/db_xref="GI:2105483"
/translation="MDVSLQDAELRDLKSHGVSGPIVATRKLYEKKLKLSGDS
INQSNLNDQSFNEDSLIISSPKSPQFQVQNSAATAATSPESDSDCEESMR
YLTEEMAADRASARQAQSNKGGLGTTTFTLEFVIAVFAYFLIENAEQLKLVAT
NPEDTI"
gene
7523..7901
/gene="nlp-12"
/note="For a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=M01D7.5;class=Seque
nce]"
CDS
join(7523..7577,7657..7715,7773..7901)
/gene="nlp-12"
/note="C. elegans putative neuropeptide preproprotein
(NLP-12); coded for by the following C. elegans cDNAs:
yk162h11.3, yk162h11.5, yk276h6.3, yk276h6.5, yk1188d11.3,
yk550c1.5"
/codon_start=1
/product="C. elegans NLP-12 protein (corresponding
sequence M01D7.5)"
/protein_id="AAB58066.1"
/db_xref="GI:2105484"
/translation="MLRHSCALLMLILVFEVFATQSPFDRQDRDYRPLQFGRKD
YRPLQFGRDYRPLQFGRKSSGSGPVLEPIWEMQ"
gene
9337..14518
/gene="M01D7.4"
/note="for a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=M01D7.8;class=Seque
nce]"
Alignment Scores:
Pred. No.: 2,6 Length: 41179
Score: 66.00 Matches: 11
Percent Similarity: 68.75% Conservative: 0
Best Local Similarity: 68.75% Mismatches: 5
Query Match: 67.35% Indels: 0
DB: 3 Gaps: 0
US-09-854-133-587 (1-16) x AF003739 (1-41179)
QY 1 PheGlnAlaAsnCysGlylleasPhePhelellePhePheTrp 16
Db 8533 TTTTCGGCAAAATGTGATTGTTTATTATTATTTTGGAAATTTGG 8580
RESULT 6
AC006895 201746 bp DNA linear HTG 26-FEB-1999
LOCUS Caenorhabditis elegans clone Y71G12, *** SEQUENCING IN PROGRESS
DEFINITION
***, 12 unordered pieces.
ACCESSION AC006895
VERSION AC006895.2 GI:4309904
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 201746)
AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
JOURNAL Waterston R.H.
TITLE The sequence of Caenorhabditis elegans clone
REFERENCE 2 (bases 1 to 201746)
AUTHORS Unpublished
JOURNAL Waterston R.H.
TITLE Direct Submission
REFERENCE 3 (bases 1 to 201746)
AUTHORS Submitted (24-FEB-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Mar 1, 1999 this sequence version replaced gi:4263453.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 2190: contig of 2190 bp in length
* 2191: contig of unknown length
* 2205: gap of unknown length
* 4263: contig of 2058 bp in length
* 4264: gap of unknown length
* 4279: contig of 2368 bp in length
* 6647: gap of unknown length
* 6662: contig of 3525 bp in length
* 10187: gap of unknown length
* 10202: contig of 3997 bp in length
* 14199: gap of unknown length
* 14213: contig of 4019 bp in length
* 18233: gap of unknown length
* 18248: contig of 5430 bp in length
* 23678: gap of unknown length
* 23693: contig of 5533 bp in length
* 29226: gap of unknown length
* 29241: contig of 8148 bp in length
* 37389: gap of unknown length
* 37404: contig of 32459 bp in length
* 69863: gap of unknown length
* 69878: contig of 5893 bp in length
* 128771: gap of unknown length
* 128786: contig of 72961 bp in length.
FEATURES
    source
    1..201746
    /organism="Caenorhabditis elegans"
    /db_xref="taxon:6239"
    /clone="Y71G12"
BASE COUNT 63828 a 37036 c 36424 g 64293 t 165 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 15 9 Length: 201746
Score: 66.00 Matches: 11
Percent Similarity: 68.75% Conservative: 0
Best Local Similarity: 68.75% Mismatches: 5
Query Match: 67.35% Indels: 5
DB: 2 Gaps: 0

```

US-09-854-133-587 (1-16) x AC008563 (1-201746)

```

QY 1 PheGlnAlaAsnCysGlyLeaSpPheIlePheTrpIlePheTrp 16
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84071 TTTTCGGCAATGTGATTTCTTTTATTTTGGAAATTTGG 84118

```

```

RESULT 7
AX251448
LOCUS
DEFINITION
Sequence 416 from Patent WO0168912.
ACCESSION
AX251448
VERSION
AX251448.1 GI:15984871
KEYWORDS
synthetic construct.
SOURCE
artificial sequences.
ORGANISM
artificial sequences.
REFERENCE
1 (bases 1 to 24259)
AUTHORS
Olek.A., Piepenbrock,C. and Berlin,K.
TITLE
Diagnosis of diseases associated with tumor suppressor genes and
conocenes
JOURNAL
Patent: WO 0168912-A 416 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
    Location/Qualifiers
    1..24259
    /organism="synthetic construct"
    /db_xref="taxon:32630"
    /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 6444 a 375 c 5691 g 11749 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 10 5 Length: 24259
Score: 61.00 Matches: 8
Percent Similarity: 90.91% Conservative: 2

```

```

Best Local Similarity: 72.73% Mismatches: 1
Query Match: 62.24% Indels: 0
DB: 6 Gaps: 0

```

US-09-854-133-587 (1-16) x AX251448 (1-24259)

```

QY 6 GlytLeaSpPheIlePheTrpIlePheTrp 16
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21896 GGTATTGATTTTAGTATTGCTTTTGGT 21928

```

RESULT 8  
AC008563/c

```

LOCUS
DEFINITION
Homo sapiens chromosome 5 clone CTC-536A23, WORKING DRAFT SEQUENCE,
7 ordered pieces.
ACCESSION
AC008563
VERSION
HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 94703)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 94703)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 6, 2000 this sequence version replaced gi:7708951.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 391027, H433
Center clone name: CIT-HSPC_536A23
-----
Summary Statistics
Consensus quality: 88918 bases at least Q40
Consensus quality: 93000 bases at least Q30
Consensus quality: 93760 bases at least Q20
Estimated insert size: 99000; pulse field gel estimation
Estimated insert size: 94453; sum-of-contigs estimation
Quality coverage: 5.84 in Q20 bases; pulse field gel estimation
Quality coverage: 6.12 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 34594: contig of 34594 bp in length
* 34595 34694: gap of unknown length
* 34695 51143: contig of 16449 bp in length
* 51144 51243: gap of unknown length
* 51244 66425: contig of 15182 bp in length
* 66426 66525: gap of unknown length
* 66526 72366: contig of 5841 bp in length
* 72367 72466: gap of unknown length
* 72467 79629: contig of 7163 bp in length
* 79630 79729: gap of unknown length
* 79730 80679: contig of 950 bp in length
* 80680 80779: gap of unknown length
* 80780 94703: contig of 13924 bp in length.
* Location/Qualifiers
* 1..94703

```

FEATURES  
source  
1..94703

ACCESSION	AC113545
VERSION	AC113545.1
KEYWORDS	HTG; HTGS_PHASE0.
SOURCE	Mus musculus.
ORGANISM	Mus musculus. Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 67679)  
Birren, B., Linton, L., Nussbaum, C. and Lander, E.  
Mus musculus, clone RP23-241C6  
Unpublished  
2 (bases 1 to 67679)  
Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choe, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,  
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,  
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,  
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE  
JOURNAL  
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L24237  
Center clone name: 241\_C\_6  
-----

\* NOTE: This record contains 82 individual  
\* sequencing reads that have not been assembled  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
704 803: contig of 703 bp in length  
804 803: gap of 100 bp  
804 1495: contig of 692 bp in length  
1496 1595: gap of 100 bp in length  
1596 2332: contig of 737 bp in length  
2333 2432: gap of 100 bp  
2433 3158: contig of 726 bp in length  
3159 3258: gap of 100 bp  
3259 3968: contig of 710 bp in length  
3969 4068: gap of 100 bp  
4069 4760: contig of 692 bp in length  
4761 4860: gap of 100 bp  
4861 5578: contig of 718 bp in length  
5579 5678: gap of 100 bp  
5679 6403: contig of 725 bp in length  
6404 6503: gap of 100 bp  
6504 7236: contig of 733 bp in length  
7237 7336: gap of 100 bp  
7337 8074: contig of 738 bp in length

8075 8174: gap of 100 bp  
8175 8880: contig of 706 bp in length  
8881 8980: gap of 100 bp  
8981 9708: contig of 728 bp in length  
9709 9808: gap of 100 bp  
9809 10544: contig of 736 bp in length  
10545 10644: gap of 100 bp in length  
10645 11350: contig of 706 bp in length  
11351 11450: gap of 100 bp  
11451 12165: contig of 715 bp in length  
12166 12265: gap of 100 bp  
12266 12984: contig of 719 bp in length  
12985 13084: gap of 100 bp  
13085 13817: contig of 733 bp in length  
13818 13917: gap of 100 bp  
13918 14663: contig of 746 bp in length  
14664 14763: gap of 100 bp  
14764 15513: contig of 750 bp in length  
15514 15613: gap of 100 bp  
15614 16344: contig of 731 bp in length  
16345 16444: gap of 100 bp  
16445 17153: contig of 709 bp in length  
17154 17253: gap of 100 bp  
17254 17993: contig of 740 bp in length  
17994 18093: gap of 100 bp  
18094 18837: contig of 744 bp in length  
18838 18937: gap of 100 bp  
18938 19685: contig of 748 bp in length  
19686 19785: gap of 100 bp  
19786 20501: contig of 716 bp in length  
20502 20601: gap of 100 bp  
20602 21311: contig of 710 bp in length  
21312 21411: gap of 100 bp  
21412 22130: contig of 719 bp in length  
22131 22230: gap of 100 bp  
22231 22944: contig of 714 bp in length  
22945 23044: gap of 100 bp  
23045 23782: contig of 738 bp in length  
23783 23882: gap of 100 bp  
23883 24623: contig of 741 bp in length  
24624 24723: gap of 100 bp  
24724 25469: contig of 746 bp in length  
25470 25569: gap of 100 bp  
25570 26275: contig of 706 bp in length  
26276 26375: gap of 100 bp  
26376 27120: contig of 745 bp in length  
27121 27220: gap of 100 bp  
27221 27906: contig of 686 bp in length  
27907 28006: gap of 100 bp  
28007 28732: contig of 726 bp in length  
28733 28832: gap of 100 bp  
28833 29544: contig of 712 bp in length  
29545 29644: gap of 100 bp  
29645 30372: contig of 728 bp in length  
30373 30472: gap of 100 bp  
30473 31190: contig of 718 bp in length  
31191 31290: gap of 100 bp  
31291 32018: contig of 728 bp in length  
32019 32118: gap of 100 bp  
32119 32871: contig of 753 bp in length  
32872 32971: gap of 100 bp  
32972 33702: contig of 731 bp in length  
33703 33802: gap of 100 bp  
33803 34551: contig of 749 bp in length  
34552 34651: gap of 100 bp  
34652 35355: contig of 704 bp in length  
35356 35455: gap of 100 bp  
35456 36199: contig of 744 bp in length  
36200 36299: gap of 100 bp  
36300 37014: contig of 715 bp in length  
37015 37114: gap of 100 bp  
37115 37824: contig of 710 bp in length  
37825 37924: gap of 100 bp

```

* 37925 38637: contig of 713 bp in length
* 38638 38737: gap of 100 bp
* 38738 39421: contig of 684 bp in length
* 39422 39521: gap of 100 bp
* 39522 40255: contig of 734 bp in length
* 40256 40355: gap of 100 bp
* 40356 41119: contig of 764 bp in length
* 41120 41219: gap of 100 bp
* 41220 41933: contig of 714 bp in length
* 41934 42033: gap of 100 bp
* 42034 42766: contig of 733 bp in length
* 42767 42866: gap of 100 bp
* 42867 43599: contig of 733 bp in length
* 43600 43699: gap of 100 bp
* 43700 44428: contig of 729 bp in length
* 44429 44528: gap of 100 bp
* 44529 45245: contig of 717 bp in length
* 45246 45345: gap of 100 bp
* 45346 46066: contig of 721 bp in length
* 46067 46166: gap of 100 bp
* 46167 46889: contig of 723 bp in length
* 46890 46989: gap of 100 bp
* 46990 47715: contig of 726 bp in length
* 47716 47815: gap of 100 bp
* 47816 48569: contig of 754 bp in length
* 48570 48669: gap of 100 bp
* 48670 49396: contig of 727 bp in length
* 49397 49496: gap of 100 bp
* 49497 50223: contig of 727 bp in length
* 50224 50323: gap of 100 bp
* 50324 51034: contig of 711 bp in length
* 51035 51134: gap of 100 bp
* 51135 51883: contig of 749 bp in length
* 51884 51983: gap of 100 bp
* 51984 52732: contig of 749 bp in length
* 52733 52832: gap of 100 bp
* 52833 53558: contig of 726 bp in length
* 53559 53658: gap of 100 bp
* 53659 54380: contig of 722 bp in length
* 54381 54480: gap of 100 bp
* 54481 55198: contig of 718 bp in length
* 55199 55298: gap of 100 bp
* 55299 56015: contig of 717 bp in length
* 56016 56115: gap of 100 bp
* 56116 56842: contig of 727 bp in length
* 56843 56942: gap of 100 bp

```

```

Alignment Scores:
Pred. No.: 75.8 Length: 67679
Score: 59.00 Matches: 7
Percent Similarity: 90.91% Conservative: 3
Best Local Similarity: 63.64% Mismatches: 1
Query Match: 60.20% Indels: 0
DB: 2 Gaps: 0

```

US-09-854-133-587 (1-16) x AC113545 (1-67679)

```

QY 6 GlylaspheillellepHeTrpIlePheTrp 16
||||:|||||: |||||:|||||
Db 26971 GCCTGGATTGTTTGTGTTTGGGTTTTTG 26939

```

```

RESULT 12
AC091629 161334 bp DNA linear ROD 19-DEC-2001
LOCUS Mus musculus chromosome 11 map 11, 64CM, Band E1, complete
DEFINITION
AC091629
ACCESSION AC091629
VERSION AC091629.1 GI:14029058
KEYWORDS HTG.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT

```

```

1 (bases 1 to 161334)
Poorkaj,P., Kas,A., D'Souza,I., Zhou,Y., Pham,Q., Stone,M.,
Olson,M.V. and Schellenberg,G.D.
A genomic sequence analysis of the mouse and human
microtubule-associated protein tau
Mamm. Genome 12 (9), 700-712 (2001)
21521499
11641718
Contact: Gerald D. Schellenberg (zachdad@u.washington.edu)
2 (bases 1 to 161334)
Kas,A.
Direct Submission
Submitted (12-MAY-2001) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
Contact: Arnold Kas(akas@u.washington.edu)
Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below
(only fragments from insert sequence is displayed). Small
fragments below a variable cutoff (approximately 400-600 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

```

BgIII			EcoRI			NsiI		
FP	Seq	FP	Seq	FP	Seq	FP	Seq	FP
1991.00	2013.00	2272.00	2258.00	4046.00	4032.00			
1788.00	1820.00	2854.00	2843.00	6952.00	6857.00			
3992.00	4005.00	9569.00	9597.00	506.00	508.00			
2149.00	2138.00	2099.00	2105.00	5420.00	5320.00			
1788.00	1753.00	1599.00	1581.00	1242.00	1233.00			
1879.00	1869.00	7895.00	7867.00	25755.00	22305.00			
1226.00	1216.00	8085.00	7933.00	7924.00	7633.00			
5423.00	5449.00	18024.00	17543.00	13876.00	13549.00			
7084.00	6935.00	828.00	801.00	7361.00	7144.00			
2654.00	2591.00	4216.00	4115.00	3148.00	3089.00			
2084.00	2056.00	5017.00	4935.00	2756.00	2736.00			
2654.00	2586.00	11890.00	11718.00	7924.00	7789.00			
985.00	970.00	3794.00	3736.00	5248.00	5145.00			
9434.00	9229.00	5245.00	5170.00	3324.00	3278.00			
869.00	851.00	4608.00	4523.00	6952.00	6725.00			
4250.00	4176.00	4216.00	4164.00	3823.00	3755.00			
1100.00	1081.00	1898.00	1875.00	1666.00	1678.00			
1541.00	1525.00	2828.00	2786.00	6602.00	6496.00			
4250.00	4230.00	625.00	613.00	10116.00	9829.00			
3992.00	3884.00	9258.00	9151.00	9151.00	8941.00			
3544.00	3448.00	1453.00	1455.00	3324.00	3257.00			
1226.00	1198.00	2416.00	2363.00					
1991.00	1950.00	1311.00	1307.00					

-----  
 5963.00 5822.00 4331.00 4286.00  
 2084.00 2107.00 1453.00 1394.00  
 869.00 850.00 3208.00 3152.00  
 1610.00 1594.00 1102.00 1090.00  
 1461.00 1442.00 2334.00 2332.00  
 2811.00 2734.00 1898.00 1900.00  
 573.00 566.00 14838.00 14692.00  
 4559.00 4489.00 6683.00 6584.00  
 3208.00 3170.00 1453.00 1443.00  
 1788.00 1723.00 3208.00 3153.00  
 1461.00 1452.00 916.00 899.00  
 4559.00 4483.00 2957.00 2900.00  
 2254.00 2208.00  
 1788.00 1789.00  
 7270.00 7150.00  
 4834.00 4757.00  
 3992.00 3956.00  
 1461.00 1474.00  
 1395.00 1392.00  
 2654.00 2587.00  
 16421.00 16316.00  
 1879.00 1857.00  
 6214.00 6135.00  
 -----

FEATURES  
 source  
 Location/Qualifiers  
 1. 161334  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="11"  
 /map="11, 64cM, Band E1"  
 /cell\_lines="Embryonic stem cell 129/sv1"  
 /clone\_lib="Genome Systems ES Release 1"  
 BASE COUNT 42340 a 37230 c 38866 g 42898 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 204 Length: 161334  
 Score: 59.00 Matches: 7  
 Percent Similarity: 90.91% Conservative: 3  
 Best Local Similarity: 63.64% Mismatches: 1  
 Query Match: 60.20% Indels: 0  
 DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AC091629 (1-161334)  
 QY 6 GlyileAspPheIlePheTrpIlePheTrp 16  
 Db 36533 GCCTTGATTTTGTGTTTGGGTTTGTG 36565  
 RESULT 13

4

AL593843  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

AL593843 198281 bp DNA linear ROD 22-MAY-2002  
 Mouse DNA sequence from clone RP23-136D4 on chromosome 11, complete  
 sequence.  
 AL593843  
 HTG.  
 AL593843.9 GI:21211822  
 house mouse.  
 Mus musculus  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Peck, A.  
 Direct Submission  
 Submitted (22-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk  
 On May 25, 2002 this sequence version replaced gi:16944181.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: En: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-136D4 is  
 from the RPCI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6.

FEATURES  
 source

Location/Qualifiers  
 1. 198281  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="11"  
 /clone="RP23-136D4"  
 /clone\_lib="RPCI-23"  
 BASE COUNT 53943 a 41053 c 44095 g 59190 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 259 Length: 198281  
 Score: 59.00 Matches: 7  
 Percent Similarity: 90.91% Conservative: 3  
 Best Local Similarity: 63.64% Mismatches: 1  
 Query Match: 60.20% Indels: 0  
 DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AL593843 (1-198281)  
 QY 6 GlyileAspPheIlePheTrpIlePheTrp 16  
 Db 130121 GCCTTGATTTTGTGTTTGGGTTTGTG 130153

RESULT 14  
 AL845270/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 AL845270 218188 bp DNA linear HTG 17-AUG-2002  
 Mus musculus chromosome 2 clone RP23-186C16, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, in ordered pieces.  
 AL845270  
 AL845270.4 GI:22416234  
 HTG; HTGS\_PHASE2; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
 house mouse.

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 218188)
AUTHORS Dunn,M.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerv@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Aug 21, 2002 this sequence version replaced gi:222655588.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerv@sanger.ac.uk
----- Project Information
Center project name: BM186C16
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 99% of reads
Consensus quality: 217952 bases at least Q40
Consensus quality: 218070 bases at least Q30
Consensus quality: 218157 bases at least Q20
Insert size: 218188; sum-of-contigs
Insert size: 201825; 1.3% error; agarose-fp
Quality coverage: 11.45x in Q20 bases; sum-of-contigs Quality
coverage: 12.57x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
Location/Qualifiers
1..218188
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-186C16"
/clone_lib="RPCI-23"
/misc_feature 1..218188
/note="assembly_fragment:04163" 1 others
BASE COUNT 67758 a 41084 c 40053 g 59292 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 288 Length: 218188
Score: 59.00 Matches: 9
Percent Similarity: 91.67% Conservative: 2
Best Local Similarity: 75.00% Mismatches: 1
Query Match: 60.20% Indels: 0
DB: 2 Gaps: 0
US-09-854-133-587 (1-16) x AL845270 (1-218188)
QY 4 AsnCysGlyIleAspPheIleIlePheTrpIlePhe 15
:::||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 80996 CATTGCAATATCGACTTTATATATATTTGGCTATTT 80961
RESULT 15
AC004984/c AC004984 149765 bp DNA linear PRI 21-DEC-1999
LOCUS Homo sapiens PAC clone RP5-1164K10 from 7p21-p22, complete
DEFINITION sequence.
ACCESSION AC004984
VERSION AC004984.1 GI:3355524
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149765)
AUTHORS Holmes,A., Elliott,G. and Biewald,T.

```

```

TITLE The sequence of Homo sapiens PAC clone RP5-1164K10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149765)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 149765)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 149765)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jul 30, 1998 this sequence version replaced gi:3213013.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplense@watson.wustl.edu
----- Summary Statistics
Center project name: H_DJ1164K10
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.
This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.
MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc
SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by
Pieter de Jong and coworkers at Roswell Park Cancer Institute,
using the method described by Ioannou et al., Nature Genetics
6:84-9 (1994). The library is from one male donor. For further
details, see http://bacpac.med.buffalo.edu/
The clone is available from Genome Systems, Inc.
(http://www.genomesystems.com).
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP5-953F6. Actual start of this
clone is at base position 1 of RP5-1164K10; actual end is at 149765
of RP5-1164K10.
Location/Qualifiers
1..149765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p21-p22"
/clone="RP5-1164K10"
/clone_lib="RPCI-5"
800..1099
/rpt_family="Alu"
repeat_region
FEATURES
source
Location/Qualifiers
1..149765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p21-p22"
/clone="RP5-1164K10"
/clone_lib="RPCI-5"
800..1099
/rpt_family="Alu"

```

```
repeat_region 1159. .1367 /rpt_family="MERL_type"
repeat_region 1522. .2141 /rpt_family="MERL_type"
repeat_region 4257. .4555 /rpt_family="Alu"
repeat_region 5498. .5793 /rpt_family="Alu"
repeat_region 5976. .6010 /rpt_family="Alu"
repeat_region 7559. .7838 /rpt_family="MERL_type"
repeat_region 7869. .7908 /rpt_family="Retroviral"
repeat_region 7909. .8211 /rpt_family="Retroviral"
repeat_region 8212. .8574 /rpt_family="Alu"
repeat_region 8851. .9024 /rpt_family="Retroviral"
repeat_region 9084. .9288 /rpt_family="MERL_type"
repeat_region 9138. .9295 /rpt_family="MERL_type"
repeat_region 9303. .9383 /rpt_family="MERL_type"
repeat_region 9582. .10468 /rpt_family="MERL_type"
repeat_region 10469. .10727 /rpt_family="MERL_type"
repeat_region 11285. .11446 /rpt_family="MERL_type"
repeat_region 11556. .11616 /rpt_family="L2"
repeat_region 11933. .12180 /rpt_family="MERL_type?"
repeat_region 12623. .12969 /rpt_family="MERL_type"
repeat_region 13221. .13348 /rpt_family="MaLR"
repeat_region 13949. .14093 /rpt_family="L2"
repeat_region 14651. .14867 /rpt_family="L2"
repeat_region 14943. .15069 /rpt_family="L2"
repeat_region 15381. .15518 /rpt_family="L2"
repeat_region 17345. .17685 /rpt_family="L2"
repeat_region 17719. .17994 /rpt_family="MERL_type"
repeat_region 18973. .19281 /rpt_family="Alu"
repeat_region 20216. .20342 /rpt_family="Alu"
repeat_region 20478. .20748 /rpt_family="MERL_type"
repeat_region 20817. .21122 /rpt_family="MERL_type"
repeat_region 21183. .21465 /rpt_family="Alu"
repeat_region 21466. .21769 /rpt_family="MERL_type"
repeat_region 21786. .22056 /rpt_family="Alu"
repeat_region 22144. .22284 /rpt_family="MaLR"
repeat_region 22294. .22546 /rpt_family="MERL_type"
repeat_region 26976. .27281 /rpt_family="MERL_type"
repeat_region 29024. .29089 /rpt_family="Alu"
```

```
repeat_region /rpt_family="MERL_type"
repeat_region 29322. .29619 /rpt_family="Alu"
repeat_region 33089. .34067 /rpt_family="L1"
repeat_region 34037. .36649 /rpt_family="L1"
STS complement(36940. .37258)
repeat_region /db_xref="GI:1185354"
repeat_region 37403. .37491 /rpt_family="L2"
repeat_region 38538. .38572 /rpt_family="MERL_type"
repeat_region 40350. .40770 /rpt_family="MaLR"
repeat_region 41667. .41912 /rpt_family="MERL_type"
repeat_region 41902. .41973 /rpt_family="MERL_type"
repeat_region 43348. .43391 /rpt_family="MaLR"
repeat_region 43455. .43692 /rpt_family="MaLR"
repeat_region 44189. .44400 /rpt_family="L1"
repeat_region 44397. .44606 /rpt_family="L1"
repeat_region 44607. .44892 /rpt_family="L1"
repeat_region 44893. .45706 /rpt_family="L1"
repeat_region 45707. .46031 /rpt_family="Alu"
repeat_region 46032. .46161 /rpt_family="L1"
repeat_region 46162. .46469 /rpt_family="Alu"
repeat_region 46470. .47221 /rpt_family="L1"
repeat_region 47222. .47523 /rpt_family="Alu"
repeat_region 47524. .47745
```

## Alignment Scores:

Pred. No.:	419	Length:	149765
Score:	57.00	Matches:	7
Percent Similarity:	83.33%	Conservative:	3
Best Local Similarity:	58.33%	Mismatches:	2
Query Match:	58.16%	Indels:	0
DB:	9	Gaps:	0

US-09-854-133-587 (1-16) x AC004984 (1-149765)

Qy 5 CysGlyIleAspPheIleIlePheTrpIlePheTrp 16

Db 58017 TGCTCTATGATGCCATCTCTTTCTGGGCTTTTGG 57982

Search completed: May 11, 2003, 15:01:07  
Job time : 461.761 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 12:41:29 ; Search time 34.6903 Seconds  
(without alignments)  
1038.677 Million cell updates/sec

Title: US-09-854-133-587

Perfect score: 98  
Sequence: 1 PQANGIDFIIFWIFW 16

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09854133/runat\_05052003\_173955\_327/app\_query.fasta\_1.462  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133@cgn2\_1.1.275 @runat\_05052003\_173955\_327 -NCPU=6 -ICPU=3  
-NO\_XLIFY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	100.0	337	23	AAD23462
2	98	100.0	2239	23	AAD23460
3	98	100.0	5981	23	AAD23461
4	61	62.2	24259	22	AAS46692
5	56	57.1	31882	23	ABL09072
6	55.5	56.6	6526	24	ABN80307
7	55.5	56.6	6526	24	ABL34237
8	55	56.1	5575	24	AAS61206
9	53	54.1	97835	24	ABK84796
10	52	53.1	740	22	AAH73248
11	52	53.1	3171	23	ABL19819
12	52	53.1	3430	23	ABL16345
13	52	53.1	6089	24	ABL32703
14	52	53.1	6285	24	ABL33497
15	52	53.1	8164	23	ABL19818
16	52	53.1	11322	23	ABL16344
17	52	53.1	611590	21	AAF22303
18	51	52.0	10957	24	ABL33111
19	51	52.0	24939	24	ABL70369
20	50.5	51.5	180	8	AAH70765
21	50.5	51.5	6713	22	AAK84891
22	50	51.0	2000	22	AAS32537
23	50	51.0	2584	19	AAV40502
24	50	51.0	8889	24	ABL32210
25	50	51.0	8925	22	ABL07244
26	50	51.0	8925	23	ABL98792
27	50	51.0	14245	21	AAZ45259
28	50	51.0	14245	21	AAZ30356
29	50	51.0	15416	24	ABL70498
30	50	51.0	15416	24	ABL34231
31	50	51.0	15416	24	ABL61453
32	50	51.0	15416	24	ABK31527
33	50	51.0	15732	22	AAS45389
34	50	51.0	15732	24	ABK28234
35	50	51.0	40324	24	ABO67149
36	49	50.0	344	22	AAH1960
37	49	50.0	582	24	ABQ26448
38	49	50.0	582	24	ABQ26449
39	49	50.0	583	24	ABQ22078
40	49	50.0	583	24	ABQ22079
41	49	50.0	4579	23	ABL05382
42	49	50.0	5143	24	ABL34567
43	49	50.0	5659	24	ABL32797
44	49	50.0	6113	24	ABL32802
45	49	50.0	17389	24	ABL33415

## ALIGNMENTS

## RESULT 1

AAD23462

ID AAD23462 standard; cdna; 337 BP.

AC AAD23462;

DT 26-FEB-2002 (first entry)

DE Human lung tumour-specific 20E10 5' cdna.

KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
antisense-therapy; vaccine; immune response; lung cancer; 20E10; ss.

XX Homo sapiens.

OS WO200172295-A2.

PN 04-OCT-2001.

PD

```

XX 28-MAR-2001; 2001WO-US09991.
PF 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX WPI; 2001-639201/73.
XX
XX New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
PI Claim 1; Page 334; 378pp; English.
XX
XX The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a cDNA encoding human lung tumour-specific protein.
XX
SQ Sequence 337 BP; 103 A; 60 C; 93 G; 81 T; 0 other;

Alignment Scores:
Pred. No.: 1.67e-06 Length: 337
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x AAD23462 (1-337)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 107 TTCAGGCCAATTGTGGCATAGATTATCATATTCGATTTTGG 154

RESULT 2
AAD23460
ID AAD23460 standard; cDNA; 2239 BP.
XX
XX AAD23460;
XX
XX 26-FEB-2002 (first entry)
XX
XX Human lung tumour-specific 19A4 cDNA.
XX
XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer; 19A4; ss.
XX
XX Homo sapiens.
XX
XX WO200172295-A2.
XX
XX 04-OCT-2001.
XX
XX 28-MAR-2001; 2001WO-US09991.
XX
XX 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX WPI; 2001-639201/73.
XX
XX New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
PI Claim 1; Page 334; 378pp; English.
XX
XX The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a cDNA encoding human lung tumour-specific protein.
XX
SQ Sequence 337 BP; 103 A; 60 C; 93 G; 81 T; 0 other;

Alignment Scores:
Pred. No.: 1.67e-06 Length: 337
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x AAD23462 (1-337)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 107 TTCAGGCCAATTGTGGCATAGATTATCATATTCGATTTTGG 154

RESULT 2
AAD23460
ID AAD23460 standard; cDNA; 2239 BP.
XX
XX AAD23460;
XX
XX 26-FEB-2002 (first entry)
XX
XX Human lung tumour-specific 19A4 cDNA.
XX
XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer; 19A4; ss.
XX
XX Homo sapiens.
XX
XX WO200172295-A2.
XX
XX 04-OCT-2001.
XX
XX 28-MAR-2001; 2001WO-US09991.
XX
XX 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.

```

```

PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX WPI; 2001-639201/73.
XX
XX New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
PI Claim 1; Page 332; 378pp; English.
XX
XX The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a cDNA encoding human lung tumour-specific protein.
XX
SQ Sequence 2239 BP; 619 A; 444 C; 493 G; 683 T; 0 other;

Alignment Scores:
Pred. No.: 1.32e-05 Length: 2239
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x AAD23460 (1-2239)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
Db 104 TTCAGGCCAATTGTGGCATAGATTATCATATTCGATTTTGG 151

RESULT 3
AAD23461
ID AAD23461 standard; cDNA; 5981 BP.
XX
XX AAD23461;
XX
XX 26-FEB-2002 (first entry)
XX
XX Human lung tumour-specific 14F10 full length cDNA.
XX
XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer; 14F10; ss.
XX
XX Homo sapiens.
XX
XX WO200172295-A2.
XX
XX 04-OCT-2001.
XX
XX 28-MAR-2001; 2001WO-US09991.
XX
XX 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;

```

XX DR WPI; 2001-639201/73.

XX PT New human lung-specific polynucleotides and polypeptides for the

XX PT diagnosis and treatment of disease e.g. lung cancer -

XX PS Claim 1; Page 332-334; 378pp; English.

XX CC The invention relates to isolated lung tumour-specific proteins and

XX CC their corresponding cDNA molecules. Lung tumour-specific proteins and

XX CC their antigen-presenting cells are useful for stimulating and/or

XX CC expanding T cells specific for a tumour protein, and for inhibiting

XX CC the development of cancer. The invention also relates to a composition

XX CC useful for stimulating an immune response, and for treating cancer. The

XX CC lung tumour specific oligonucleotide is useful in gene therapy and for

XX CC diagnosis, detection and treatment of lung cancer. The present sequence

XX CC is a cDNA encoding human lung tumour-specific protein.

XX SQ Sequence 5981 BP; 1858 A; 1029 C; 1098 G; 1996 T; 0 other;

Alignment Scores:

Pred. No.:	3.88e-05	Length:	5981
Score:	98.00	Matches:	16
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	23	Gaps:	0

US-09-854-133-587 (1-16) x AAD23461 (1-5981)

Qy 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpPheTrp 16

Db 102 TTCAGGCCAATGTCGACATATTTATCATATTCGATTTTGG 149

RESULT 4

AAS46692

ID AAS46692 standard; DNA; 24259 BP.

XX AC AAS46692;

XX DT 18-DEC-2001 (first entry)

XX TTumour suppressor gene derived chemically modified sequence #415.

XX DE Human; tumour suppressor gene; oncogene; antitumour; cytostatic;

XX KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;

XX KW cytosine methylation; ds.

XX OS Homo sapiens.

XX PN WO200168912-A2.

XX PD 20-SEP-2001.

XX PF 15-MAR-2001; 2001WO-EP02955.

XX PR 15-MAR-2000; 2000DE-1013847.

XX PR 06-APR-2000; 2000DE-1019058.

XX PR 07-APR-2000; 2000DE-1019173.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2001-602752/68.

XX PT Fragments of chemically modified genes associated with tumour suppressor

XX PT genes and oncogenes, useful in designing primers and probes for

XX PT analysing diseases associated with cytosine methylation state e.g.

XX PT cancer -

PS Claim 1; SEQ ID NO 415; 27pp; English.

XX CC The invention relates to a nucleic acid comprising a sequence of 18

XX CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

XX CC bisulphite, of genes associated with tumour suppression and

XX CC oncogenes having a sequence taken from 536 (actually 533 since

XX CC numbers 408, 458 and 500 are missing from the sequence listing) sequences

XX CC (SS) and sequences complementary to (SS). The nucleic acid may be a

XX CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may

XX CC form part of a set of probes for detecting the cytosine methylation state

XX CC and/or single nucleotide polymorphisms and also to be used in an

XX CC array for analysing diseases associated with CpG dinucleotides e.g.

XX CC cancers and tumours. The probes can also be used in a method for

XX CC ascertaining genetic and/or epigenetic parameters for the diagnosis

XX CC and/or therapy of existing diseases or the predisposition to specific

XX CC diseases, by analysing cytosine methylations. The parameters may be

XX CC compared to another set of genetic and/or epigenetic parameters, the

XX CC differences serving as basis for diagnosis and/or prognosis events which

XX CC are disadvantageous to patients. The present sequence is one of the

XX CC 533 genomic sequences derived from tumour suppressor genes and

XX CC oncogenes.

XX CC Note: The sequence data for this patent did not form part

XX CC of the printed specification, but was obtained in electronic

XX CC format directly from WIPO at

XX CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 24259 BP; 6444 A; 375 C; 5691 G; 11749 T; 0 other;

Alignment Scores:

Pred. No.:	88.5	Length:	24259
Score:	61.00	Matches:	8
Percent Similarity:	90.91%	Conservative:	2
Best Local Similarity:	72.73%	Mismatches:	1
Query Match:	62.24%	Indels:	0
DB:	22	Gaps:	0

US-09-854-133-587 (1-16) x AAS46692 (1-24259)

Qy 6 GlyIleAspPheIleIlePheTrpPheTrp 16

Db 21896 GGTATTGATTTTATAGTATTGTTTGG 21928

RESULT 5

ABL09072/c

ID ABL09072 standard; cDNA; 31882 BP.

XX AC ABL09072;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 21698.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR P-PSDB; ABB64969.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX interactions -

PS Claim 1; SEQ ID NO 21698; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 31882 BP; 9206 A; 6377 C; 6605 G; 9694 T; 0 other;

Alignment Scores:  
 Pred. No.: 701 Length: 31882  
 Score: 56.00 Matches: 7  
 Percent Similarity: 76.92% Conservative: 3  
 Best Local Similarity: 53.85% Mismatches: 3  
 Query Match: 57.14% Indels: 0  
 DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABL09072 (1-31882)

QY 4 AsnCysGlyLeaspPheIlePheTrpIlePheTrp 16

DB 27609 AATTGGGGTAGCTTTTGTTCCTTGGCTATTGG 27571

RESULT 6

ABL80307

ID ABL80307 standard; DNA; 6526 BP.

XX AC ABL80307;

XX DT 15-JUL-2002 (first entry)

XX DE Human chemically modified disease associated gene SEQ ID NO 324.

XX KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
 KW antidiabetic; cytostatic; anticonvulsant; ds.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200200927-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP07536.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-130908/17.

XX Novel nucleic acid useful for diagnosis and therapy of diseases  
 XX associated with development genes such as diabetes, comprises a  
 PT sequence of a segment of chemically pretreated DNA of genes associated  
 PT with development -

XX Claim 1; SEQ ID NO 324; 27pp; English.

XX

CC The invention relates to a nucleic acid (I) comprising a sequence at  
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
 CC of genes associated with development selected from 87 genes listed in  
 CC the specification such as ACCPN, ADFN, or AFD1 and comprising one of 350  
 CC sequences (ABN7984-ABN80333) or their complements. The invention is  
 CC useful for the diagnosis or therapy of diseases associated with  
 CC development genes, in particular disease related to homeobox containing  
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Currarino syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
 CC Oligomers specific to each of the genes are useful for detecting the  
 CC methylation state of all CpG dinucleotides within the 350 sequences or  
 CC (II) and their complementary sequences, as primer oligonucleotides for  
 CC the amplification of the 350 sequences, (II) and/or their complements and  
 CC as oligomer probes for detecting the cytosine methylation state and/or  
 CC single nucleotide polymorphisms (SNPs).

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.

XX SQ Sequence 6526 BP; 1009 A; 556 C; 2240 G; 2721 T; 0 other;

Alignment Scores:

Pred. No.: 148 Length: 6526  
 Score: 55.50 Matches: 9  
 Percent Similarity: 73.33% Conservative: 2  
 Best Local Similarity: 60.00% Mismatches: 1  
 Query Match: 56.63% Indels: 3  
 DB: 24 Gaps: 1

US-09-854-133-587 (1-16) x ABLN80307 (1-6526)

QY 5 CysGlyLeaspPheIlePheTrp-----IlePheTrp 16

DB 1151 TGTGGAATGTTTATAGTTTTTGTGTTTACGCTTATTGG 1195

RESULT 7

ABL34237

ID ABL34237 standard; DNA; 6526 BP.

XX AC ABL34237;

XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 2210.

XX KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

XX OS Homo sapiens.

XX PN WO200200928-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP07537.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-130909/17.

Tue May 13 12:13:01 2003

us-09-854-133-587.rng

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation

XX Claim 1; SEQ ID NO 2210; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX Sequence 6526 BP; 1009 A; 556 C; 2240 G; 2721 T; 0 other;

Alignment Scores:  
Pred. No.: 148 Length: 6526  
Score: 55.50 Matches: 9  
Percent Similarity: 73.33% Conservative: 2  
Best Local Similarity: 60.00% Mismatches: 1  
Query Match: 56.63% Indels: 3  
DB: 24 Gaps: 1

US-09-854-133-587 (1-16) x ABL34237 (1-6526)

QY 5 CysGlytLeAspPheIleIlePheTrp-----IlePheTrp 16  
DB 1151 TGTGGAATGGTTTATAGTTTGTGTTTACGTTATTTGG 1195

RESULT 8  
ID AAS61206  
XX AAS61206 standard; DNA: 5575 BP.  
XX AAS61206;  
XX 29-JAN-2002 (first entry)  
XX Human gene regulation-associated gene oligonucleotide #161.  
XX Human: Gene regulation-associated gene; severe combined immunodeficiency;  
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
KW renal disease; Freecampia; cardiac allograft vascular disease;  
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
KW immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;  
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
XX Homo sapiens.  
OS WO200177375-A2.  
XX 18-OCT-2001.  
XX 06-APR-2001; 2001WO-EP03968.  
XX 06-APR-2000; 2000DE-1019058.  
XX 07-APR-2000; 2000DE-1019173.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
PI WPI; 2002-017470/02.  
XX New nucleic acid sequences from chemically modified genes associated  
PT with gene regulation, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
PT disease

XX Claim 1; SEQ ID NO 165; 26pp; English.

XX The invention relates to 224 nucleic acid sequences comprising at least  
CC 18 bases of a chemically pretreated gene associated with gene regulation  
CC selected from 43 known genes (or complementary sequences). The  
CC chemical pretreatment converts cytosine bases unmethylated at the  
CC 5-position to uracil or another base with hybridisation behaviour  
CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
CC The DNA sequences, oligomers (or sets/arrays) and method are  
CC useful in the diagnosis of disease and in therapy of such diseases,  
CC associated with gene regulation and in therapy of such diseases, by  
CC enabling analysis of the cytosine methylation patterns of such genes,  
CC kits are provided. They are especially useful in diagnosis  
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
CC preclampsia, graft versus-host disease. The present sequence is a  
CC sequence included in the sequence data for this specification and is  
CC associated with the human gene regulation-associated genes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 5575 BP; 1212 A; 190 C; 1534 G; 2639 T; 0 other;

Alignment Scores:  
Pred. No.: 149 Length: 5575  
Score: 55.00 Matches: 8  
Percent Similarity: 80.00% Conservative: 4  
Best Local Similarity: 53.33% Mismatches: 3  
Query Match: 56.12% Indels: 0  
DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x AAS61206 (1-5575)

QY 1 PheGlnAlaAsnCysGlytLeAspPheIleIlePheTrpIlePhe 15  
DB 1697 TTGAGGAATTCGGGTAGATTGTTGATTTTATTTT 1741

RESULT 9  
ABR84796/C  
ID ABR84796 standard; CDNA: 97835 BP.  
XX ABR84796;  
XX 14-AUG-2002 (first entry)  
XX Human cDNA differentially expressed in granulocytic cells #1367.  
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX Homo sapiens.  
OS WO200228999-A2.  
XX 11-APR-2002.  
XX 03-OCT-2001; 2001WO-US30821.  
XX 03-OCT-2000; 2000US-237189P.  
XX (GENE-) GENE LOGIC INC.  
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
PI

XX WPI; 2002-435328/46.  
 XX Detecting granulocyte activation by detecting differential expression  
 PT of genes associated with granulocyte activation, which serves as  
 PT diagnostic markers that is useful for monitoring disease states and  
 PT drug toxicity -  
 XX  
 PS Claim 1; SEQ ID No 1367; 114pp; English.  
 XX  
 CC The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing  
 CC the expression level to an expression level in an unactivated  
 CC GC, where differential expression of Gs is indicating of GCA.  
 CC Also included are modulating (M2) GA by contacting GC with an agent  
 CC that alters the expression of Gs.  
 CC for an agent capable of modulating GCA or an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease using the  
 CC gene expression profile; (3) detecting (M4) an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease, by detecting the  
 CC level of expression in a sample of the tissue of gene(s) from Gs, where  
 CC the level of expression of the gene is indicative of inflammation;  
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
 CC an allergic response in a subject, exposure of a subject to a pathogen  
 CC or sterile inflammatory disease, by contacting a tissue having  
 CC inflammation with an agent that modulates the expression of gene(s)  
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
 CC GCA preferably in an inflammation in a tissue; M4 is useful for  
 CC detecting an inflammation (especially chronic) in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
 CC periodontal disease; also bacterial infection, viral infection, M5 is  
 CC parasitic infection, protozoal infection, fungal infection, and M5 is  
 CC useful for treating one of the above conditions. The present  
 CC sequence represents a gene differentially expressed in granulocytes.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 97835 BP; 26355 A; 21843 C; 22546 G; 27091 T; 0 other;

Alignment Scores:  
 Pred. No.: 6.92e+03 Length: 97835  
 Score: 53.00 Matches: 7  
 Percent Similarity: 90.91% Conservative: 3  
 Best Local Similarity: 63.64% Mismatches: 1  
 Query Match: 54.08% Indels: 0  
 DB: 24 Gaps: 0

US-09-854-133-587 (1-16) x ABK84796 (1-97835)  
 QY 5 CysGlyleasppheillellePheTrpIlePhe 15  
 ID AAH73248/c  
 Db 68056 TGTGGAATTGAATTTTAAATTTTATGGTGT 68024  
 RESULT 10  
 AAH73248/c  
 ID AAH73248 standard; cDNA; 740 BP.  
 XX  
 AC AAH73248;  
 XX  
 DT 19-SEP-2001 (first entry)  
 DE Human cervical cancer marker nucleic acid 4522.  
 XX

XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
 XX Homo sapiens.  
 XX WO200142467-A2.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 08-DEC-2000; 2000WO-US33312.  
 XX  
 PR 08-DEC-1999; 99US-0169681.  
 PR 21-DEC-1999; 99US-0171350.  
 PR 14-MAR-2000; 2000US-0189315.  
 PR 12-MAY-2000; 2000US-0203791.  
 PR 09-JUN-2000; 2000US-0210600.  
 PR 21-JUL-2000; 2000US-0220114.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Deeds J, Berger A, Zhao X;  
 XX WPI; 2001-375006/39.  
 DR  
 XX  
 PT New isolated nucleic acid for diagnosing and treating cervical cancer  
 XX and for assessing and detecting compounds for treating the cancer -  
 PS Claim 1; Page 994; 1051pp; English.  
 XX  
 CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
 CC polypeptides are useful: to assess if a patient is afflicted with  
 CC cervical cancer or has a pre-malignant condition; to monitor the  
 CC progression of cervical cancer or a premalignant condition in a patient;  
 CC and to select and/or assess the efficacy of a compound or therapy for  
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
 CC useful for gene therapy.  
 XX  
 SQ Sequence 740 BP; 292 A; 129 C; 127 G; 177 T; 15 other;

Alignment Scores:  
 Pred. No.: 47.3 Length: 740  
 Score: 52.00 Matches: 8  
 Percent Similarity: 81.82% Conservative: 1  
 Best Local Similarity: 72.73% Mismatches: 2  
 Query Match: 53.06% Indels: 0  
 DB: 22 Gaps: 0

US-09-854-133-587 (1-16) x AAH73248 (1-740)  
 QY 6 GlytleasppheillellePheTrpIlePheTrp 16  
 ID ABL19819/c  
 Db 714 GGATGCTTTTAAATATTTTGGATCTTTGG 682  
 RESULT 11  
 ABL19819/c  
 ID ABL19819 standard; DNA; 3171 BP.  
 XX  
 AC ABL19819;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 10930.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX interactions -  
 XX Claim 1; SEQ ID NO 10930; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 XX capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX useful in developmental biology and in elucidating cell signalling and  
 XX cell-cell interactions in higher eukaryotes for the development of  
 XX insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 XX sequences (ABL01840-ABL16175) and the encoded proteins  
 XX (ABB57737-ABB72072).  
 XX The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 3171 BP; 955 A; 806 C; 740 G; 670 T; 0 other;

Alignment Scores:  
 Pred. No.: 232 Length: 3171  
 Score: 52.00 Matches: 8  
 Percent Similarity: 62.50% Conservative: 2  
 Best Local Similarity: 50.00% Mismatches: 6  
 Query Match: 53.06% Indels: 0  
 DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABL19819 (1-3171)  
 QY 1 PheGlnAlaAsnCysGlyLeuAspPheIleIlePheTrpIlePheTrp 16  
 DB 284 TTCCTCGCTGGACATGGATTCGTTGTGTGCTTCTGCTCTCTGG 237

RESULT 12  
 ABL16345/c  
 ID ABL16345 standard; DNA; 3430 BP.  
 XX ABL16345;  
 AC  
 XX 26-MAR-2002 (first entry)  
 DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 508.  
 XX Drosophila melanogaster genomic polynucleotide; cell signalling; insecticide;  
 KW Drosophila; developmental biology; gene; ds.  
 KW pharmaceutical; gene; ds.  
 XX Drosophila melanogaster.  
 OS WO200171042-A2.  
 PN 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US09231.  
 PF 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX Claim 1; SEQ ID NO 508; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 XX capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX useful in developmental biology and in elucidating cell signalling and  
 XX cell-cell interactions in higher eukaryotes for the development of  
 XX insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 XX sequences (ABL01840-ABL16175) and the encoded proteins  
 XX (ABB57737-ABB72072).  
 XX The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 3430 BP; 1028 A; 859 C; 812 G; 731 T; 0 other;

Alignment Scores:  
 Pred. No.: 253 Length: 3430  
 Score: 52.00 Matches: 8  
 Percent Similarity: 62.50% Conservative: 2  
 Best Local Similarity: 50.00% Mismatches: 6  
 Query Match: 53.06% Indels: 0  
 DB: 23 Gaps: 0

US-09-854-133-587 (1-16) x ABL16345 (1-3430)  
 QY 1 PheGlnAlaAsnCysGlyLeuAspPheIleIlePheTrpIlePheTrp 16  
 DB 543 TTCCTCGCTGGACATGGATTCGTTGTGTGCTTCTGCTCTCTGG 496

RESULT 13  
 ABL32703  
 ID ABL32703 standard; DNA; 6089 BP.  
 XX ABL32703;  
 AC  
 XX 26-MAR-2002 (first entry)  
 DT Human immune system associated gene SEQ ID NO: 676.  
 XX Human immune system associated gene  
 DE Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianemic; cytosolic; neurotropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX Homo sapiens.  
 OS WO200200928-A2.  
 PN 03-JAN-2002.  
 PD 02-JUL-2001; 2001WO-EP07537.  
 PF 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX (EPIG-) EPIGENOMICS AG.  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI WPI; 2002-130909/17.  
 DR Nucleic acid comprising fragment of chemically modified gene, useful  
 XX for diagnosis and treatment of diseases associated with abnormal  
 PT

PT cytosine methylation -

XX Claim 1; SEQ ID NO 676; 32pp + Sequence Listing; German.

PS The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX Sequence 6089 BP; 1370 A; 240 C; 1823 G; 2656 T; 0 Other;

SQ

Alignment Scores:

Pred. No.:	474	Length:	6089
Score:	52.00	Matches:	7
Percent Similarity:	90.91%	Conservative:	3
Best Local Similarity:	63.64%	Mismatches:	1
Query Match:	53.06%	Indels:	0
DB:	24	Gaps:	0

US-09-854-133-587 (1-16) x ABL32703 (1-6089)

QY 5 CysGlyIleAspPheIleIlePheTrpIlePhe 15

DB 4865 TGTGGTATTGAGTATTGATTTTGGTATTTT 4897

RESULT 14

ABL33497

ID ABL33497 standard; DNA; 6285 BP.

XX ABL33497;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1470.

XX Human; immune system disease; cytosine methylation; antiasthmatic;

XX antiarteriosclerotic; antiandemic; cytostatic; nootropic;

XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

XX antineumatic; antiarthritis; antidiabetic; antipsoriatic;

XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

XX gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful

XX for diagnosis and treatment of diseases associated with abnormal

XX cytosine methylation -

XX Claim 1; SEQ ID NO 1470; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

XX genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX Sequence 6285 BP; 1346 A; 263 C; 1715 G; 2961 T; 0 Other;

SQ

Alignment Scores:

Pred. No.:	490	Length:	6285
Score:	52.00	Matches:	9
Percent Similarity:	69.23%	Conservative:	0
Best Local Similarity:	69.23%	Mismatches:	4
Query Match:	53.06%	Indels:	0
DB:	24	Gaps:	0

US-09-854-133-587 (1-16) x ABL33497 (1-6285)

QY 4 AsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16

DB 4522 AATTAGGGAATAGGATTTATTTTGGGATTTTGG 4560

RESULT 15

ABL19818

ID ABL19818 standard; DNA; 8164 BP.

XX ABL19818;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 10927.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 10927; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 8164 BP; 2296 A; 1589 C; 1831 G; 2448 T; 0 Other;



Tue May 13 12:13:01 2003

Alignment Scores: 653 Length: 8164  
Pred. No.: 52.00 Matches: 8  
Score: 62.50% Conservative: 2  
Percent Similarity: 50.00% Mismatches: 6  
Best Local Similarity: 53.06% Indels: 0  
Query Match: 23 Gaps: 0  
DB: 23  
US-09-854-133-587 (1-16) x ABL19818 (1-8164)  
QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
Db 6218 TTCTCGCTGGACATGATTCGTCCTTTGTGTGCTTCGGCTCTCTGG 6265

Search completed: May 11, 2003, 14:16:32  
Job time : 49.6903 secs





## RESULT 2

US-09-780-049-18  
; Sequence 18, Application US/09780049  
; Patent No. 6465250  
; GENERAL INFORMATION:  
; APPLICANT: Brett P. Monia  
; APPLICANT: Jacqueline Whatt  
; TITLE OF INVENTION: ANTISENSE MODULATION OF PROTEIN PHOSPHATASE 2 CATALYTIC SUBUNIT  
; TITLE OF INVENTION: EXPRESSION  
; FILE REFERENCE: RTS-0134  
; CURRENT APPLICATION NUMBER: US/09/780,049  
; CURRENT FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 96  
; SEQ ID NO 18  
; LENGTH: 40000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-780-049-18

## Alignment Scores:

Pred. No.:	322	Length:	40000
Score:	53.00	Matches:	7
Percent Similarity:	80.00%	Conservative:	5
Best Local Similarity:	46.67%	Mismatches:	3
Query Match:	54.08%	Indels:	0
DB:	4	Gaps:	0

US-09-854-133-587 (1-16) x US-09-780-049-18 (1-40000)

QY 2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16

Db 16687 CAAACACAGTGTGACTAATACATCTCTTTTCTTTTGG 16731

## RESULT 3

US-07-638-431-1/c  
; Sequence 1, Application US/07638431  
; Patent No. 5198535  
; GENERAL INFORMATION:  
; APPLICANT: Hoffman, Stephen L.  
; APPLICANT: Charoenvit, Yupin  
; APPLICANT: Hedstrom, Richard  
; APPLICANT: Khushmith, Srisin  
; APPLICANT: Rogers IV, William O.  
; TITLE OF INVENTION: Protective malaria sporozoite surface protein  
; TITLE OF INVENTION: immunogen and gene  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: A. David Spevack  
; STREET: NMRDC Building 1 T-12 National Naval  
; CITY: Bethesda  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20814-5044  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/638,431  
; FILING DATE: 19910110  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, Avram D.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-4033  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4673 base pairs  
; TYPE: NUCLEIC ACID

; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: N  
; ANTI-SENSE: N  
; ORIGINAL SOURCE:  
; ORGANISM: Plasmodium yoelii  
; STRAIN: 17X(NL)  
; DEVELOPMENTAL STAGE: erythrocytic stage  
; TISSUE TYPE: Blood  
; CELL TYPE: erythrocytic stage  
; IMMEDIATE SOURCE:  
; LIBRARY: Py-lambdagtl1-2-7 kb genomic expression  
; CLONE: Py10.1111  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 718..3195  
; OTHER INFORMATION:  
US-07-638-431-1

Alignment Scores:			
Pred. No.:	184	Length:	4673
Score:	48.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	48.98%	Indels:	0
DB:	1	Gaps:	0

US-09-854-133-587 (1-16) x US-07-638-431-1 (1-4673)

QY 9 PheIleIlePheTrpIlePheTrp 16

Db 2601 TTCAATATCTCTGGGTTTATTGG 2578

## RESULT 4

PCT-US92-00018-1/c  
; Sequence 1, Application PC/TUS9200018  
; GENERAL INFORMATION:  
; APPLICANT: Hoffman, Stephen L.  
; APPLICANT: Charoenvit, Yupin  
; APPLICANT: Hedstrom, Richard  
; APPLICANT: Khushmith, Srisin  
; APPLICANT: Rogers IV, William O.  
; TITLE OF INVENTION: Protective malaria sporozoite surface protein  
; TITLE OF INVENTION: immunogen and gene encoding  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: A. David Spevack  
; STREET: NMRDC Building 1 T-12 National Naval  
; CITY: Bethesda  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20814-5044  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00018  
; FILING DATE: 19920103  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, Avram D.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-4033  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4673 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double

```

; FILING DATE: 8-SEPTEMBER-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/931.473
; FILING DATE: 17-AUGUST-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX 34.2/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA
; FEATURE:
; OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-046-247-138

Alignment Scores:
Pred. No.: 3,19 Length: 66
Score: 46.00 Matches: 7
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 58.33% Mismatches: 4
Query Match: 46.94% Indels: 0
DB: 3 Gaps: 0

US-09-854-133-587 (1-16) x US-09-046-247-138 (1-66)

QY 5 CysGlyIleAspPheIlellePheTrpIlePheTrp 16
|||||
Db 11 UCGGUUAGCUCUUUGUUUUUUUUUUUUUUUUUGCCGG 46

RESULT 6
US-08-981-527A-20
; Sequence 20, Application US/08981527A
; Patent No. 6410262
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Kerkman, Richard
; TITLE OF INVENTION: Broekhuizen, Cornelis P.
; TITLE OF INVENTION: No. 6410262el Secretion Factors for E
; TITLE OF INVENTION: Gram-Positive Microorganisms Genes E
; TITLE OF INVENTION: of Using It
; FILE REFERENCE: GCX322-US
; CURRENT APPLICATION NUMBER: US/08/981,527A
; CURRENT FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: PCT/NL96/00278
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 4370
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-08-981-527A-20

Alignment Scores:
Pred. No.: 360 Length: 4370
Score: 46.00 Matches: 8
Percent Similarity: 75.00% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 4
Query Match: 46.94% Indels: 0
DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-08-981-527A-20 (1-4370)

QY 1 PheclnAlaasnCysGlyIleAspPheIlellePheTrpIlePheTrp 1
|||||
Db 3839 TTCATATACAAATTCGGAATAAATGGATGATATTATTTAGGTATTTTGG 1
```

```

RESULT 7
US-08-545-196B-20/c
; Sequence 20, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 18..881
US-08-545-196B-20
Alignment Scores:
Pred. No.: 86.4 Length: 885
Score: 45.00 Matches: 10
Percent Similarity: 46.43% Conservatives: 3
Best Local Similarity: 35.71% Mismatches: 3
Query Match: 45.92% Indels: 12
DB: 3 Gaps: 2
US-09-854-133-587 (1-16) x US-08-545-196B-20 (1-885)
QY 1 PheGlnAlaAnCysGlyle-----AspPheIlelle----- 11
Db 278 TTTCAAGGAGTGTGGCATCTCTTTGGCTTTATCTCTTGGCAGGTTTCTCT 219
QY 12 -----PheTrpIlePheTrp 16
Db 218 GGCTGTGCCTTTTGGCTTATCTGTG 195
RESULT 8
US-08-257-073-14/c
; Sequence 14, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisné, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE

```

```

; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-257-073-14
Alignment Scores:
Pred. No.: 183 Length: 1725
Score: 45.00 Matches: 6
Percent Similarity: 100.00% Conservatives: 1
Best Local Similarity: 85.71% Mismatches: 0
Query Match: 45.92% Indels: 0
DB: 1 Gaps: 0
US-09-854-133-587 (1-16) x US-08-257-073-14 (1-1725)
QY 10 IleIlePheTrpIlePheTrp 16
Db 1194 ATCGTCTCTGGATTTTGG 1174
RESULT 9
US-09-484-970B-134
; Sequence 134, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 134
; LENGTH: 2435
; TYPE: DNA

```

ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 OTHER INFORMATION: Incyte ID No. 6426186 247789.2CB1  
 NAME/KEY: unsure  
 LOCATION: 93, 128, 132, 143-144, 2419, 2427, 2429  
 OTHER INFORMATION: a, t, c, g, or other  
 US-09-484-970B-134

Alignment Scores:  
 Pred. No.: 270 Length: 2435  
 Score: 45.00 Matches: 6  
 Percent Similarity: 90.00% Conservative: 3  
 Best Local Similarity: 60.00% Mismatches: 1  
 Query Match: 45.92% Indels: 0  
 DB: 4 Gaps: 0

US-09-854-133-587 (1-16) x US-09-484-970B-134 (1-2435)

QY 6 GlylleAspPheIlePheTrrPilePhe 15  
 |||||::: ::|||:::||||  
 Db 1253 GGGATCGACTGGTCTCTTCGGGCTTC 1282

RESULT 10  
 US-08-920-812-10/c  
 ; Sequence 10, Application US/08920812  
 ; Patent No. 5763188  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ohno, Tsuneya  
 ; APPLICANT: Matsuhisa, Akio  
 ; APPLICANT: Uehara, Hirotsugu  
 ; APPLICANT: Eda, Soji  
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/920,812  
 ; FILING DATE: 29-AUG-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/362,577  
 ; FILING DATE: 27-MAR-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rin-Laures, Li-Hsien  
 ; REGISTRATION NUMBER: 33,547  
 ; REFERENCE/DOCKET NUMBER: 19036/32420  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3719 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Enterococcus faecalis  
 ; STRAIN: Clinical Isolate S2-3  
 ; US-08-920-812-10

Alignment Scores:  
 Pred. No.: 435 Length: 3719  
 Score: 45.00 Matches: 7  
 Percent Similarity: 54.55% Conservative: 5  
 Best Local Similarity: 31.82% Mismatches: 2  
 Query Match: 45.92% Indels: 8  
 DB: 1 Gaps: 1

US-09-854-133-587 (1-16) x US-08-920-812-10 (1-3719)

QY 1 PheGlnAlaSerCysGlylle-----AspPheIleIlePhe 12  
 |||||::: ::|||:::||||  
 Db 1878 TTTAAACGGACTGTGGTGTGTTTTCACGAACACTACCGAAGCGACTCTTTTGTGTT 1819

QY 13 TrpIle 14  
 |||:::  
 Db 1818 TGGCTT 1813

RESULT 11  
 US-08-920-827-10/c  
 ; Sequence 10, Application US/08920827  
 ; Patent No. 5770375  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ohno, Tsuneya  
 ; APPLICANT: Matsuhisa, Akio  
 ; APPLICANT: Uehara, Hirotsugu  
 ; APPLICANT: Eda, Soji  
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/920,827  
 ; FILING DATE: 29-AUG-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/362,577  
 ; FILING DATE: 27-MAR-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rin-Laures, Li-Hsien  
 ; REGISTRATION NUMBER: 33,547  
 ; REFERENCE/DOCKET NUMBER: 19036/32420  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3719 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Enterococcus faecalis  
 ; STRAIN: Clinical Isolate S2-3  
 ; US-08-920-827-10

Alignment Scores:  
 Pred. No.: 435 Length: 3719  
 Score: 45.00 Matches: 7  
 Percent Similarity: 54.55% Conservative: 5  
 Best Local Similarity: 31.82% Mismatches: 2

```

Query Match: 45.92% Indels: 8
DB: 1 Gaps: 1

US-09-854-133-587 (1-16) x US-08-920-827-10 (1-3719)

QY 1 PheGlnAlaAsnCysGlyLe-----AspPheIleIlePhe 12
Db 1878 TTTAGAACGGACTGTGTGTTTGTTCACGAACACCGGAACCGACTCTTTTTCGTTT 1819
QY 13 Ttpile 14
Db 1818 TGGCTT 1813

RESULT 12
US-08-921-177-10/c
; Sequence 10, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsumura, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3719 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecalis
; STRAIN: Clinical Isolate S2-3
; US-08-921-177-10

Alignment Scores:
Pred. No.: 435 Length: 3719
Score: 45.00 Matches: 7
Percent Similarity: 54.55% Conservative: 5
Best Local Similarity: 31.82% Mismatches: 2
Query Match: 45.92% Indels: 8
DB: 1 Gaps: 1

US-09-854-133-587 (1-16) x US-08-921-177-10 (1-3719)

```



RESULT 14

US-08-920-828-10/c  
 ; Sequence 10, Application US/08920828  
 ; Patent No. 585398  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ohno, Tsuneya  
 ; APPLICANT: Matsuhisa, Akio  
 ; APPLICANT: Uehara, Hirotsugu  
 ; APPLICANT: Eda, Soji  
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/920,828  
 ; FILING DATE: 29-AUG-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/362,577  
 ; FILING DATE: 27-MAR-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rin-Laures, Li-Hsien  
 ; REGISTRATION NUMBER: 33,547  
 ; REFERENCE/DOCKET NUMBER: 19036/32420  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3719 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Enterococcus faecalis  
 ; STRAIN: Clinical Isolate S2-3  
 ; PS-08-920-828-10

Alignment Scores:		
Pred. No.:	435	3719
Score:	45.00	Length:
Percent Similarity:	54.55%	Matches:
Best Local Similarity:	31.82%	Conservative:
Query Match:	45.92%	Mismatches:
DB:	2	Indels:
		Gaps:

US-09-854-133-587 (1-16) x US-08-920-828-10 (1-3719)

QY	1	PheClnAlaAsnCysGlyIle-----AspPheIleIlePhe	12
		:::: :::::	::::: :::::
Dd	1878	TTTAAAGCGACTGTGGTGTTTTCACGAACCTACCGAAGCGACTCTCTTTTGT	1819
QY	13	Tripile	14
Dd	1818	TGGCTT	1813

RESULT 15

US-09-596-824-5/C  
; Sequence 5, Application US/09596824  
; Patent No. 6372464  
; GENERAL INFORMATION:



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 15:28:15 ; Search time 18.9735 seconds  
(without alignments)  
1047.953 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FQANGIDFIIFWIFW 16

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q-/cgn2\_1/USPFO\_spool/US09854133/runat\_05052003\_173958\_449/app\_query.fasta\_1.462  
-DB=PublishedApplications\_NA -QPM=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USPR=US09854133 @CGN\_1\_1\_117 @runat\_05052003\_173958\_449  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	100.0	337	9 US-09-854-133-442	Sequence 442, App
2	98	100.0	337	10 US-09-738-973-442	Sequence 442, App
3	98	100.0	2239	9 US-09-854-133-440	Sequence 440, App
4	98	100.0	2239	10 US-09-738-973-440	Sequence 440, App

5	98	100.0	5981	9 US-09-854-133-441	Sequence 441, App
6	98	100.0	5981	10 US-09-738-973-441	Sequence 441, App
7	51	52.0	806	10 US-09-910-943-590	Sequence 590, App
8	51	52.0	31412	9 US-10-109-551-3	Sequence 3, Appli
9	50.5	51.5	710	10 US-09-867-550-1537	Sequence 1537, Ap
10	50	51.0	8925	9 US-09-764-891-9932	Sequence 9932, Ap
11	50	51.0	15732	9 US-10-239-676-96	Sequence 96, Appl
12	50	51.0	1691139	9 US-10-067-514-1	Sequence 1, Appli
13	49	50.0	2157	10 US-09-801-368-387	Sequence 387, App
14	49	50.0	83450	9 US-09-811-463-3	Sequence 3, Appli
15	48	49.0	303	10 US-09-998-598-2471	Sequence 2471, Ap
16	48	49.0	405	9 US-09-918-995-33363	Sequence 33363, A
17	48	49.0	601	10 US-09-777-921A-64	Sequence 64, Appl
18	48	49.0	2285	9 US-09-978-295A-283	Sequence 283, App
19	48	49.0	2285	9 US-09-978-697-283	Sequence 283, App
20	48	49.0	2285	9 US-09-978-192A-283	Sequence 283, App
21	48	49.0	2285	9 US-09-999-832A-283	Sequence 283, App
22	48	49.0	2285	9 US-09-978-183-283	Sequence 283, App
23	48	49.0	2285	9 US-10-174-590-111	Sequence 111, App
24	48	49.0	2285	9 US-10-176-758-111	Sequence 111, App
25	48	49.0	2285	9 US-10-173-737-111	Sequence 111, App
26	48	49.0	2285	9 US-10-173-706-111	Sequence 111, App
27	48	49.0	2285	9 US-10-173-738-111	Sequence 111, App
28	48	49.0	2285	9 US-10-175-752-111	Sequence 111, App
29	48	49.0	2285	9 US-10-176-482-111	Sequence 111, App
30	48	49.0	2285	9 US-10-176-757-111	Sequence 111, App
31	48	49.0	2285	9 US-10-176-913-111	Sequence 111, App
32	48	49.0	2285	9 US-10-180-552-111	Sequence 111, App
33	48	49.0	2285	9 US-10-180-557-111	Sequence 111, App
34	48	49.0	2285	9 US-10-173-700-111	Sequence 111, App
35	48	49.0	2285	9 US-10-174-572-111	Sequence 111, App
36	48	49.0	2285	9 US-10-174-575-111	Sequence 111, App
37	48	49.0	2285	9 US-10-174-582-111	Sequence 111, App
38	48	49.0	2285	9 US-10-174-588-111	Sequence 111, App
39	48	49.0	2285	9 US-10-175-733-111	Sequence 111, App
40	48	49.0	2285	9 US-10-175-740-111	Sequence 111, App
41	48	49.0	2285	9 US-10-175-743-111	Sequence 111, App
42	48	49.0	2285	9 US-10-176-488-111	Sequence 111, App
43	48	49.0	2285	9 US-10-176-492-111	Sequence 111, App
44	48	49.0	2285	9 US-10-176-747-111	Sequence 111, App
45	48	49.0	2285	9 US-10-176-750-111	Sequence 111, App

#### ALIGNMENTS

RESULT 1  
US-09-854-133-442  
; Sequence 442, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 442  
; LENGTH: 337  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-854-133-442

Alignment Scores:  
Pred. No.: 2.04e-07 Length: 337  
Score: 98.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

```
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-854-133-442 (1-337)

Qy 1 PheGlnAlaAsnCysGlyleAspPheIleIlePheTrpIlePheTrp 16
|||||
Db 107 TTCCAGGCCAATTGTGGCATAGATTTATCATATTCGTGGATTTTGG 154

RESULT 2
US-09-738-973-442
; Sequence 442, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738, 973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 442
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-738-973-442

Alignment Scores:
Pred. No.: 2,04e-07 Length: 337
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-738-973-442 (1-337)

Qy 1 PheGlnAlaAsnCysGlyleAspPheIleIlePheTrpIlePheTrp 16
|||||
Db 107 TTCCAGGCCAATTGTGGCATAGATTTATCATATTCGTGGATTTTGG 154

RESULT 3
US-09-854-133-440
; Sequence 440, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 440
; LENGTH: 2239
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-09-854-133-440

Alignment Scores:
Pred. No.: 1,91e-06 Length: 2239
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-854-133-440 (1-2239)

Qy 1 PheGlnAlaAsnCysGlyleAspPheIleIlePheTrpIlePheTrp 16
|||||
Db 104 TTCCAGGCCAATTGTGGCATAGATTTATCATATTCGTGGATTTTGG 151

RESULT 4
US-09-738-973-440
; Sequence 440, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738, 973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 440
; LENGTH: 2239
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-738-973-440

Alignment Scores:
Pred. No.: 1,91e-06 Length: 2239
Score: 98.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-738-973-440 (1-2239)

Qy 1 PheGlnAlaAsnCysGlyleAspPheIleIlePheTrpIlePheTrp 16
|||||
Db 104 TTCCAGGCCAATTGTGGCATAGATTTATCATATTCGTGGATTTTGG 151

RESULT 5
US-09-854-133-441
; Sequence 441, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 440
; LENGTH: 2239
; TYPE: DNA
```

; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854, 133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 441  
; LENGTH: 5981  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-854-133-441

Alignment Scores:  
Pred. No.: 6.1e-06 Length: 5981  
Score: 98.00 Matches: 16  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-854-133-441 (1-5981)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
|||||  
Db 102 TTCCAGGCCAATGTGCGCATAGATTTATCATATCTCGGATTTTGG 149

## RESULT 6

US-09-738-973-441  
; Sequence 441, Application US/09738973  
; Patent No. US20020110563A1

## GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodon  
; APPLICANT: Aligate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738, 973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 441  
; LENGTH: 5981  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-738-973-441

Alignment Scores:  
Pred. No.: 6.1e-06 Length: 5981  
Score: 98.00 Matches: 16  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-738-973-441 (1-5981)

QY 1 PheGlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
|||||  
Db 102 TTCCAGGCCAATGTGCGCATAGATTTATCATATCTCGGATTTTGG 149

## RESULT 7

US-09-910-943-590/c  
; Sequence 590, Application US/09910943  
; Patent No. US20020081610A1

## GENERAL INFORMATION:

; APPLICANT: Hemmati-Brivanlou, Ali  
; APPLICANT: Altman, Curtis  
; TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression  
; FILE REFERENCE: 7529/1G148US1  
; CURRENT APPLICATION NUMBER: US/09/910,943  
; CURRENT FILING DATE: 2001-07-23  
; NUMBER OF SEQ ID NOS: 742  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 590  
; LENGTH: 806  
; TYPE: DNA  
; ORGANISM: Xenopus laevis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(806)  
; OTHER INFORMATION: n may be a or g or c or t/u  
US-09-910-943-590

## Alignment Scores:

Pred. No.: 19.9 Length: 806  
Score: 51.00 Matches: 6  
Percent Similarity: 75.00% Conservatives: 3  
Best Local Similarity: 50.00% Mismatches: 3  
Query Match: 52.04% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x US-09-910-943-590 (1-806)

QY 5 CysGlyIleAspPheIleIlePheTrpIlePheTrp 16  
|||||  
Db 154 TGTGTATCTATCATGTGCTCTCTGCTGGTCTCTGG 119

## RESULT 8

US-10-109-551-3/c  
; Sequence 3, Application US/10109551  
; Publication No. US20020194635A1

## GENERAL INFORMATION:

; APPLICANT: DUNNE, PATRICK W.  
; APPLICANT: PIEDRAHITA, JORGE  
; TITLE OF INVENTION: TRANSGENIC ANIMALS RESISTANT TO TRANSMISSIBLE  
; TITLE OF INVENTION: SPONGIFORM ENCEPHALOPATHIES  
; FILE REFERENCE: TAMK:207US  
; CURRENT APPLICATION NUMBER: US/10/109,551  
; CURRENT FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: 60/280,549  
; PRIOR FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 31412  
; TYPE: DNA  
; ORGANISM: Ovis aries  
US-10-109-551-3

## Alignment Scores:

Pred. No.: 1.5e+03 Length: 31412  
Score: 51.00 Matches: 6  
Percent Similarity: 90.00% Conservatives: 3  
Best Local Similarity: 60.00% Mismatches: 1  
Query Match: 52.04% Indels: 0  
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-109-551-3 (1-31412)

QY 7 IleAspPheIleIlePheTrpIlePheTrp 16  
:::|||||  
Db 15883 CTAGATTTGTTTATCTGGATCTTTGG 15854

## RESULT 9

US-09-867-550-1537  
; Sequence 1537, Application US/09867550  
; Patent No. US20020082206A1

```
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1537
; LENGTH: 710
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-550-1537

Alignment Scores:
Pred. No.: 20.6 Length: 710
Score: 50.50 Matches: 8
Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 3
Query Match: 51.53% Indels: 1
DB: 10 Gaps: 1

US-09-854-133-587 (1-16) x US-09-867-550-1537 (1-710)
QY 4 AsnCysGly---IleAspPheIleIlePheTrpIlePheTrp 16
   ||| ||| :: |||||::||| |||||
Db 174 AACGCTGCAGGTTAGTTTCATTTTGTGTTTTTTGG 215

RESULT 10
US-09-764-891-9932/G
; Sequence 9932, Application US/09/764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9932
; LENGTH: 8925
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9932

Alignment Scores:
Pred. No.: 492 Length: 8925
Score: 50.00 Matches: 8
Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 4
Query Match: 51.02% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-09-764-891-9932 (1-8925)
QY 3 AlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
   |||::| ||||| |||||::|||
Db 5144 GCATCACTCGATAGACAGATACATCTTCGGCTATCTGG 5103

RESULT 11
US-10-239-676-96
; Sequence 96, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 96
; LENGTH: 15732
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-96

Alignment Scores:
Pred. No.: 962 Length: 15732
Score: 50.00 Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 51.02% Indels: 0
DB: 9 Gaps: 0

US-09-854-133-587 (1-16) x US-10-239-676-96 (1-15732)
QY 4 AsnCysGlyIleAspPheIleIlePheTrpIlePhe 15
   ::|||::|::||| ||||| |||
Db 475 AGTTGTGAATGATTTAGTTTGTGTTTTTTT 510

RESULT 12
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretaardt, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reynisdottir, Sigridur Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

Alignment Scores:
Pred. No.: 2.24e+05 Length: 1691139
Score: 50.00 Matches: 7
Percent Similarity: 73.33% Conservative: 4
Best Local Similarity: 46.67% Mismatches: 2
Query Match: 51.02% Indels: 2
DB: 9 Gaps: 1

US-09-854-133-587 (1-16) x US-10-067-514-1 (1-1691139)
QY 4 AsnCysGlyIle-----AspPheIleIlePheTrpIlePheTrp 16
```

\_\_\_\_\_





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 11, 2003, 14:07:50 ; Search time 242.549 seconds  
(without alignments)  
1068.354 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FOANGCIDFIPIWIFW 16

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q/cn2\_1/USPTO\_spool/US09854133/runat\_05052003\_173956\_349/app\_query.fasta\_1.462  
-DB-EST -QFMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854133 @CN\_1\_1.2013 @runat\_05052003\_173956\_349 -NCPUP=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:.\*  
1: em\_estba:.\*  
2: em\_esthum:.\*  
3: em\_estin:.\*  
4: em\_estmu:.\*  
5: em\_estov:.\*  
6: em\_estpl:.\*  
7: em\_estro:.\*  
8: em\_htc:.\*  
9: gb\_est1:.\*  
10: gb\_est2:.\*  
11: gb\_hic:.\*  
12: gb\_est3:.\*  
13: gb\_est4:.\*  
14: gb\_est5:.\*  
15: em\_estfun:.\*  
16: em\_estom:.\*  
17: gb\_gss:.\*  
18: em\_gss\_hum:.\*  
19: em\_gss\_inv:.\*  
20: em\_gss\_pln:.\*  
21: em\_gss\_vrt:.\*  
22: em\_gss\_fun:.\*  
23: em\_gss\_man:.\*  
24: em\_gss\_mus:.\*  
25: em\_gss\_other:.\*  
26: em\_gss\_pro:.\*  
27: em\_gss\_rod:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
C 1	59	60.2	813	17	BH437997 BOHJ144TR
C 2	57	58.2	322	9	AV169412 AV169412
C 3	57	58.2	402	10	AW845948 PM4-CT015
C 4	57	58.2	402	10	AW845949 PM4-CT015
C 5	57	58.2	650	17	AZ340196 IN0072H24
C 6	57	58.2	751	12	BF274788 GA_ED002
C 7	57	58.2	1231	13	BI085630 602870174
C 8	56	57.1	173	12	BF471824 UI-M-BH3
C 9	56	57.1	239	12	BE948302 UI-M-BH3
C 10	56	57.1	327	12	BF225111 u243b01.y
C 11	56	57.1	337	12	BF320777 u256b08.y
C 12	56	57.1	383	12	BE986193 UI-M-CG0P
C 13	56	57.1	467	9	AA967175 ua50c04.r
C 14	56	57.1	470	10	BE200533 ug70c08.y
C 15	56	57.1	516	9	AI527491 ufi12c07.y
C 16	56	57.1	579	14	BQ564930 q127e11.y
C 17	56	57.1	583	14	BQ086748 h89c11.y
C 18	56	57.1	588	12	BF136680 601780445
C 19	56	57.1	599	10	AW213068 un98a09.y
C 20	56	57.1	622	14	BQ921841 AGENCOURT
C 21	56	57.1	642	10	AW989255 uf24d09.y
C 22	56	57.1	667	12	BF142738 601789992
C 23	56	57.1	726	13	BG919850 602823591
C 24	56	57.1	732	10	AW213203 uo01405.y
C 25	56	57.1	898	14	BQ878072 AGENCOURT
C 26	55	56.1	168	14	BQ375999 CM0-TN003
C 27	55	56.1	214	17	AQ005390 CIT-HSP-2
C 28	55	56.1	358	9	AI970001 wq79c06.x
C 29	55	56.1	512	17	AQ768884 HS_3191.A
C 30	55	56.1	515	10	AW510874 hc89d06.x
C 31	54	55.1	212	10	BE057036 G126 Giga
C 32	54	55.1	252	9	AV255114 AV255114
C 33	54	55.1	583	10	AV719226 AV719226
C 34	54	55.1	617	17	AZ004528 RPCI-23-3
C 35	54	55.1	820	17	BH063329 RPCI-24-3
C 36	54	55.1	1062	17	CNS051TA
C 37	53	54.1	303	17	TA212D11Q
C 38	53	54.1	481	17	BH277396
C 39	53	54.1	807	12	BG839567
C 40	53	54.1	830	13	BI889060 2F637-2-0
C 41	53	54.1	972	17	AG133150 Pan trogl
C 42	53	54.1	1012	17	AG138611 Pan trogl
C 43	52.5	53.6	257	12	BF835845 CM0-HT095
C 44	52.5	53.6	325	10	AV333550 AV333550
C 45	52.5	53.6	688	9	AI067874 EST209558

# ALIGNMENTS

RESULT 1  
BH437997/c  
LOCUS BOHJ144TR BOHJ Brassica oleracea genomic clone BOHJ144, DNA  
DEFINITION sequence.  
ACCESSION BH437997.1 GI:17623718  
VERSION BH437997.1  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica;  
1 (bases 1 to 813)  
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOHJ144TF  
Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

## FEATURES

Location/Qualifiers

1..813

/organism="Brassica oleracea"

/strain="T01000DH3"

/db\_xref="taxon:3712"

/clone\_lib="BOHJ144"

/note="Vector: pHS1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHS1 using BstXI linkers"

273 a 209 c 150 g 181 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 102 Length: 813

Score: 59.00 Matches: 8

Percent Similarity: 75.00% Conservatives: 4

Best Local Similarity: 50.00% Mismatches: 4

Query Match: 60.20% Indels: 0

DB: 17 Gaps: 0

US-09-854-133-587 (1-16) x BH437997 (1-813)

QY 1 PheGlnAlaAsnCysGlylleAspPheTrpPheTrpPheTrp 16

Db 708 TTTTATGCCATGCTAGGCTAGATTCATTATTATTTGGTTTCTGG 661

RESULT 2

AV169412/c

LOCUS

DEFINITION

AV169412

AV169412

AV169412.1 GI:5375849

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Carninci.P., Shibata.K., Ozawa.Y., Konno.H., Itoh.M., Aizawa.K.,

Akahira.S., Akiyama.J., Fukuda.S., Fukunishi.Y., Funayama.T., Hara

.K., Hayatsu.N., Hori.F., Ishikawa.T., Itoh.M., Izawa.M., Kawai.J.,

Kiuchi.N., Kojima.Y., Matsuyama.T., Niitsuma.H., Oda.H., Owa.C.,

Sato.K., Shibata.Y., Shigemoto.Y., Shiraki.T., Sogabe.Y., Sugahara

.Y., Suzuki.H., Suzuki.H., Tateo.M., Tomaru.Y., Tomioka.N.,

Watanabe.S., Yagame.M., Yamamura.T., Yokota.I., Yoshino.M.,

Yuranatsu.M., Okazaki.Y. and Hayashizaki.Y.

RIKEN Mouse ESTs

Unpublished (1999)

Contact: Chile Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@tcr.riken.go.jp

Thermolabile enzymes by

thermolabile and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES Location/Qualifiers

1..322

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="3110084I03"

/clone\_lib="Mus musculus head C57BL/6J 13-day embryo"

/sex="mixed"

/tissue\_type="head"

/dev\_stage="13-day embryo"

/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia

) with a modified polylinker; Site\_1: Not I; Site\_2: Eco

RI; 1st strand cDNA was primed with a Not I - oligo(dT)

primer [5'

TGTTACCAATCTGAAGTGGGAGCGCGCAATGTTTTTTTTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M.Fatima

Bonaldo."

BASE COUNT 110 a 81 c 31 g 100 t

ORIGIN

Alignment Scores:

Pred. No.: 81.6 Length: 322

Score: 57.00 Matches: 7

Percent Similarity: 77.78% Conservatives: 7

Best Local Similarity: 38.89% Mismatches: 0

Query Match: 58.16% Indels: 4

DB: 9 Gaps: 1

US-09-854-133-587 (1-16) x AV169412 (1-322)

QY 3 AlaAsnCysGly-----IleAspPheIlePheTrpPheTrp 16

Db 157 AGCTCTGCGGCTGGTGTAGTTGAATTTTGGTGTGGTTATTTGG 104

RESULT 3

AW845948/c

LOCUS

DEFINITION

PM4-CT0152-220999-001-A05 CT0152 Homo sapiens cDNA, mRNA sequence.

AV845948

AV845948.1 GI:7941465

EST.

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 402)

Dias Neto.E., Garcia Correa.R., Verjovski-Almeida.S., Briones.M.R.,

Naqai.M.A., da Silva.W. Jr., Zago.M.A., Bordin.S., Costa.F.F.,

Goldman.G.H., Carvalho.A.F., Matsukuma.A., Bata.G.S., Simpson.D.H.,

Brunstein.A., deoliveira.P.S., Bucher.P., Jongeneel.C.V., O'Hare

.M.J., Soares.F., Brentani.R.R., Reis.L.F., de Souza.S.J. and

Simpson.A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=PM4-CT0152-220>)

999-001-A05et3=1999-09-22et4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 402.

#### FEATURES

source

Location/Qualifiers  
1. .402  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0152"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 131 a 90 c 82 g 99 t  
ORIGIN

#### Alignment Scores:

Pred. No.: 100 Length: 402  
Score: 57.00 Matches: 8  
Percent Similarity: 76.92% Conservative: 2  
Best Local Similarity: 61.54% Mismatches: 3  
Query Match: 58.16% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AW845948 (1-402)

QY 4 AsnCysGlylleAspPheIlePheTrpIlePheTrp 16  
:::|||||::: |||||||  
DB 212 TCCTGCGGAATTAACCTAGCGCATTTGGATATTGG 174

#### RESULT 4

AW845949/c

LOCUS

DEFINITION PM4-CT0152-220999-001-A08 CT0152 Homo sapiens cDNA, mRNA EST 19-MAY-2000

ACCESSION AW845949

VERSION AW845949.1 GI:7941466

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 402)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-pm4-CT0152-220

999-001-A08et3=1999-09-22et4=1)

Seq primer: puc 18 forward

High quality sequence stop: 402.

Location/Qualifiers

1. .402

/organism="Homo sapiens"

/db\_xref="taxon:9606"

#### FEATURES

source

/clone\_lib="CT0152"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 131 a 90 c 82 g 99 t  
ORIGIN

#### Alignment Scores:

Pred. No.: 100 Length: 402  
Score: 57.00 Matches: 8  
Percent Similarity: 76.92% Conservative: 2  
Best Local Similarity: 61.54% Mismatches: 3  
Query Match: 58.16% Indels: 0  
DB: 10 Gaps: 0

US-09-854-133-587 (1-16) x AW845949 (1-402)

QY 4 AsnCysGlylleAspPheIlePheTrpIlePheTrp 16  
:::|||||::: |||||||  
DB 212 TCCTGCGGAATTAACCTAGCGCATTTGGATATTGG 174

#### RESULT 5

AZ340196/c

LOCUS

DEFINITION 1M0072H24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0072H24 F, DNA sequence.

ACCESSION AZ340196

VERSION AZ340196.1 GI:10415403

KEYWORDS GSS.

SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 650)

Dunn, D., Aoyagi, A., Barber, M., Bescorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0072 row: H column: 24

Seq primer: CGTTGTAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 650.

Location/Qualifiers

1. .650

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0072H24"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a







```

QY 2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
   |||
Db 289 CAGCTGACATGTCGCAAAATTTTGTGTTGTTGTTGTTGGTGTGG 333
   |||
RESULT 12
BE986193/c
LOCUS
DEFINITION
  UI-M-CG0p-bgp-b-09-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
  UI-M-CG0p-bgp-b-09-0-UI 3', mRNA sequence.
ACCESSION
  BE986193
VERSION
  BE986193.1 GI:10660263
KEYWORDS
  EST.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
  1 (bases 1 to 383)
REFERENCE
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
JOURNAL
  97044477
MEDLINE
  Contact: Chin, H
  National Institute of Mental Health
  6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
  20892-9643, USA
  Tel: 301 443 1706
  Fax: 301 443 9890
  Email: MEST@mail.nih.gov
  The sequence contained an oligo-dT track that was present in the
  oligonucleotide that was used to prime the synthesis of first
  strand cDNA and therefore this may represent a bonafide poly A
  tail. The sequence tag present in the cDNA between the NotI site
  and the oligo-dT track served to identify it as a clone from the
  retina tissue cDNA library preparation: M.B. Soares Lab Clone
  distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
  GENETICS. It should be noted that Bento Soares is generating a
  small number of additional specialized non-redundant arrays of BMAP
  cDNAs whose availability will be considered under appropriate and
  limited collaborative arrangements The tissue for this library was
  contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine
  Seq primer: M13 Forward
  POLYA=yes.
FEATURES
  source
  1..383
  /organism="Mus musculus"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="UI-M-CG0p-bgp-b-09-0-UI"
  /lab_lib="NIH_BMAP_Ret4_S2"
  /note="Vector: p7T3D-Pac (Pharmacia) with a modified
  polylinker; Site_1: Not I; Site_2: Eco RI; The
  NIH_BMAP_Ret4_S2 library is a subtracted library,
  ultimately derived from mouse retina tissue libraries at
  various stages of development. For a detailed description
  of the library from which this clone was derived, please
  visit our web site at brainest.eng.uiowa.edu. The tissue
  for this library was contributed by Dr. Xin-Yuan Fu, Yale
  University School of Medicine
  TAG_LIB=NIH_BMAP_Ret4_S2
  TAG_TISSUE=adult-retina
  TAG_SEQ=GTACGCGGCAC"
BASE COUNT 127 a 84 c 65 g 105 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 132 Length: 383
Score: 56.00 Matches: 7
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 46.67% Mismatches: 5
Query Match: 57.14% Indels: 0

```

```

DB: 12 Gaps: 0
US-09-854-133-587 (1-16) x BE986193 (1-383)
QY 2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
   |||
Db 167 CAGCTGACATGTCGCAAAATTTTGTGTTGTTGTTGTTGGTGTGG 123
   |||
RESULT 13
AA967175
LOCUS
DEFINITION
  ua50c04.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
  IMAGE:1350150 5', similar to gb:X65627 M.musculus mRNA TNZ2 for p68
  RNA helicase (MOUSE);, mRNA sequence.
ACCESSION
  AA967175
VERSION
  AA967175.1 GI:3141068
KEYWORDS
  EST.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 467)
REFERENCE
  Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
  Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
  Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
  Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
  Waterston,R.
  The WashU-HMI Mouse EST Project
  Unpublished (1996)
  Contact: Marra M/Mouse EST Project
  WashU-HMI Mouse EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@watson.wustl.edu
  This clone is available royalty-free through LLNL ; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:698942
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 385.
Location/Qualifiers
  1..467
  /organism="Mus musculus"
  /strain="C57BL/6 x CBA"
  /db_xref="taxon:10090"
  /clone="IMAGE:1350150"
  /sex="female"
  /tissue_type="lung"
  /dev_stage="6-8 month old"
  /lab_host="SOLR (kanamycin resistant)"
  /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
  ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
  dt. 6-8 month old female lung and 1.5 year old male lung
  were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP
  XR Vector: -5' adaptor sequence: 5' GAATTCGCGCAG 3' -3'
  adaptor sequence: 5' CTCGAGCTTTTGTGTTTGTGTTT 3'"
BASE COUNT 117 a 97 c 88 g 165 t
ORIGIN
Alignment Scores:
Pred. No.: 159 Length: 467
Score: 56.00 Matches: 7
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 46.67% Mismatches: 5
Query Match: 57.14% Indels: 0
DB: 9 Gaps: 0
US-09-854-133-587 (1-16) x AA967175 (1-467)
QY 2 GlnAlaAsnCysGlyIleAspPheIleIlePheTrpIlePheTrp 16
   |||

```





Result No.	Query Score	% Match	Length	DB	ID	Description
1	16	16.5	16	22	AAE13851	Human T cell epit
2	9	9.3	188	22	ABG04555	Novel human diag
3	8	8.2	46	22	AAO04640	Human polypeptide
4	8	8.2	47	22	AAE03401	Human gene 9 encod
5	8	8.2	47	23	ABG63681	Human albumin fusi
6	8	8.2	109	22	AAU32547	Novel human secret
7	7	7.2	15	22	AAAG68058	Human DNA mismatch
8	7	7.2	16	19	AAW50918	Amino acid sequenc
9	7	7.2	28	22	AAW03335	Human immune/haema
10	7	7.2	32	22	AAO05715	Human polypeptide

PIED, NO. is the number of times the score of the result being printed is greater than or equal to the score of the total score distribution.

Result No.	Query Score	% Match	Length	DB	ID	Description
1	16	16.5	16	22	AAE13851	Human T cell epit
2	9	9.3	188	22	ABG04555	Novel human diag
3	8	8.2	46	22	AAO04640	Human polypeptide
4	8	8.2	47	22	AAE03401	Human gene 9 encod
5	8	8.2	47	23	ABG63681	Human albumin fusi
6	8	8.2	109	22	AAU32547	Novel human secret
7	7	7.2	15	22	AAAG68058	Human DNA mismatch
8	7	7.2	16	19	AAW50918	Amino acid sequenc
9	7	7.2	28	22	AAW03335	Human immune/haema
10	7	7.2	32	22	AAO05715	Human polypeptide

Human T cell epitope.  
XX  
DE  
Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
antitumor therapy; vaccine; immune response; lung cancer;  
T cell epitope.  
KW  
KW

XX OS Homo sapiens.  
 XX PN WO200172295-A2.  
 XX PD 04-OCT-2001.  
 XX PF 28-MAR-2001; 2001WO-US09991.  
 XX PR 29-MAR-2000; 2000US-0538037.  
 XX PR 05-JUN-2000; 2000US-0588937.  
 XX PR 18-AUG-2000; 2000US-0640878.  
 XX PR 22-SEP-2000; 2000US-2345178.  
 XX PR 01-NOV-2000; 2000US-0704512.  
 XX PR 14-DEC-2000; 2000US-0738973.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Read SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;  
 XX PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kallos MD;  
 XX DR WPI; 2001-639201/73.  
 XX PA New human lung-specific polynucleotides and polypeptides for the  
 XX PT diagnosis and treatment of disease e.g. lung cancer  
 XX PS Claim 2; Page 378; 378pp; English.  
 XX CC The invention relates to isolated lung tumour-specific proteins and  
 XX CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 XX CC their antigen-presenting cells are useful for stimulating and/or  
 XX CC the development of cancer. The invention also relates to a composition  
 XX CC useful for stimulating an immune response, and for treating cancer. The  
 XX CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 XX CC diagnosis, detection and treatment of lung cancer. The present sequence  
 XX CC is human T cell epitope related to lung tumour-specific protein.  
 XX SQ Sequence 16 AA;  
 Query Match 16.5%; Score 16; DB 22; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-10;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 35 FOANCGIDFIIFWIFW 50  
 DB 1 FOANCGIDFIIFWIFW 16  
 RESULT 2  
 ABG04555  
 ID ABG04555 standard; Protein; 188 AA.  
 AC ABG04555;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #4546.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS68742.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 XX biodiversity  
 PS Claim 20; SEQ ID No 34914; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG03077 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 188 AA;  
 Query Match 9.3%; Score 9; DB 22; Length 188;  
 Best Local Similarity 100.0%; Pred. No. 0.22;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 26 ERKKRERK 34  
 DB 66 ERKKRERK 74  
 RESULT 3  
 AAO04640  
 ID AAO04640 standard; Protein; 46 AA.  
 AC AAO04640;  
 XX DT 06-NOV-2001 (first entry)  
 XX DE Human polypeptide SEQ ID NO 18532.  
 XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 XX KW nervous system disorders; arthritis; inflammation.  
 XX OS Homo sapiens.  
 XX PN WO200164835-A2.  
 XX PD 07-SEP-2001.  
 XX PF 26-FEB-2001; 2001WO-US04927.  
 XX PR 28-FEB-2000; 2000US-0515126.  
 XX PR 18-MAY-2000; 2000US-0577409.

us-09-854-133-586.olig.rag

Tue May 13 12:12:42 2003

```

PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
PI WPI; 2001-514838/56.
XX N-PSDB; AAI84571.
DR Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
PT Claim 20; SEQ ID NO 18532; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
PS the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and/or
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 46 AA;
SQ
Query Match 8.2%; Score 8; DB 22; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.67; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;
QY 23 RKRRKKK 30
DB 20 RKRRKKK 27
RESULT 4
AAE03401
ID AAE03401 standard; Protein; 47 AA.
XX AAE03401;
AC 10-AUG-2001 (first entry)
DT Human gene 9 encoded secreted protein HKMLP68, SEQ ID NO:59.
DE Human; secreted protein; proliferative disorder; cancer; tumour;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX inflammation; allergy; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; sepsis; cognitive disorder; schizophrenia; asthma;
XX skin disorder; psoriasis; diabetes; diabetes; atherosclerosis;
XX cardiovascular disorder; angiogenic disorder; kidney disorder;
XX gastrointestinal disorder; pregnancy-related disorder;
XX endocrine disorder; infection; wound healing; vulnery;
XX cell culture; chemotaxis; food additive; gene therapy;
XX binding partner identification.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Peptide 1
FT /Label= Signal_peptide
FT 2..47
FT Protein /Label= Mature_human_secreted_protein
ET
XX WO200132687-A1.
PN
XX 10-MAY-2001.
PD
XX

```

```

XX 25-OCT-2000; 2000WO-US29364.
XX 29-OCT-1999; 99US-0162238.
XX 30-JUN-2000; 2000US-0215134.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Komatsoulis GA, Duan RD, Fiscella M, Ebner R;
XX WPI; 2001-308740/32.
XX N-PSDB; AAD07793.
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX Claim 11; Page 376; 398pp; English.
XX AAD07770-AAD07794 represent cDNAs corresponding to 10 human secreted
XX protein genes, and AAE03378-AAE03402 represent the proteins they encode.
XX AAE03403-AAE03415 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX therapy. Pathological conditions can be diagnosed by determining the
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 10 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of
XX proliferative disorders, cancer, tumours, foetal and developmental
XX abnormalities, haematopoietic disorders, diseases of the immune system,
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX allergies, neurological disorders (e.g., Alzheimer's disease,
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX cardiovascular disorders, angiogenic disorders, kidney disorders,
XX gastrointestinal disorders, pregnancy-related disorders, endocrine
XX disorders, and infections. The proteins can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin ageing due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues, to identify their
XX cognate ligands or binding partners, and in chemotaxis, and can be used
XX as a food additive or preservative to modify storage properties.
XX Antibodies specific for a protein of the invention can be used in
XX alleviating symptoms associated with the disorders mentioned above, and
XX in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX immunosorbent assay (ELISA). The present sequence represents a human
XX secreted protein of the invention.
XX Sequence 47 AA;
SQ
Query Match 8.2%; Score 8; DB 22; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.68; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;
QY 23 RKRRKKK 30
DB 36 RKRRKKK 43
RESULT 5
ABG63681
ID ABG63681 standard; Protein; 47 AA.
XX ABG63681;
XX 27-AUG-2002 (first entry)
XX Human albumin fusion protein #356.
XX
XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
XX human serum albumin; HSA; cancer; reproductive disorder;
XX digestive disorder; immune disorder; endocrine disorder;
XX haematopoietic disorder; neural disorder; connective disorder;
XX cytostatic; antiinfertility; antiinflammatory; antiulcer;

```

KW Immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;  
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritic.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX WO200177137-A1.  
 PN 18-OCT-2001.  
 PD 12-APR-2001; 2001WO-US11988.  
 PF 12-APR-2000; 2000US-229358P.  
 XX 25-APR-2000; 2000US-199384P.  
 PR 21-DEC-2000; 2000US-256931P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Haseltine WA;  
 XX WPI; 2002-010886/01.  
 PI New fusion protein for treating disease e.g. diabetes comprises an  
 DR albumin fused to a therapeutic protein.  
 XX Claim 1; Page 789; 2102pp; English.  
 CC The present invention relates to albumin fusion proteins comprising a  
 CC therapeutic protein X and human albumin (HA), also known as human serum  
 CC albumin (HSA). The proteins are useful for treating a disease or  
 CC disorder that may be modulated by therapeutic protein X. The albumin  
 CC extends the shelf-life of protein X, and may increase its biological  
 CC in vitro/in vivo activity. The protein is useful for treating and  
 CC diagnosing disorders such as cancer, reproductive disorders, digestive  
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders  
 CC (e.g. diabetes), haematopoietic disorders, neural disorders  
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,  
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders  
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin  
 CC fusion proteins of the invention.  
 XX SQ Sequence 47 AA;  
 Query Match 8.2%; Score 8; DB 23; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 0.68;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKKK 30  
 Db 36 RKKERKKK 43  
 RESULT 6  
 AAU32547  
 ID AAU32547 standard; Protein; 109 AA.  
 AC AAU32547;  
 XX 18-DEC-2001 (first entry)  
 DT Novel human secreted protein #3038.  
 DE Human; vaccination; gene therapy; nutritional supplement;  
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 OS Homo sapiens.  
 XX WO200179449-A2.  
 XX 25-OCT-2001.  
 PD

XX 16-APR-2001; 2001WO-US08656.  
 XX 18-APR-2000; 2000US-0552929.  
 PR 26-JAN-2001; 2001US-0770160.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT;  
 PI WPI; 2001-611725/70.  
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic  
 DR vaccination, testing and therapy.  
 PT Claim 20; Page 629; 765pp; English.  
 XX The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising  
 CC the nucleic acids encoding the polypeptides and cells genetically  
 CC engineered to express them are also useful for producing the proteins.  
 CC The proteins are useful in genetic vaccination, testing and  
 CC therapy, and can be used as nutritional supplements, and in  
 CC increase stem cell proliferation; to regulate haematopoiesis; and in  
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid  
 CC sequences of novel human secreted proteins of the invention.  
 XX SQ Sequence 109 AA;  
 Query Match 8.2%; Score 8; DB 22; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 1.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 28 KKKRERKF 35  
 Db 83 KKKRERKF 90  
 RESULT 7  
 AAG68058  
 ID AAG68058 standard; Peptide; 15 AA.  
 XX AC AAG68058;  
 XX 13-DEC-2001 (first entry)  
 DT Human DNA mismatch repair protein 15 N-terminal peptide SEQ ID NO:7.  
 DE Human; DNA mismatch repair protein 15; cytostatic; virucidal;  
 KW immunomodulatory; anti-inflammatory; haemostatic; anti-HIV; gene therapy;  
 KW diagnosis; malignant tumour; haemopathy; human immunodeficiency virus;  
 KW HIV infection; immunological disease; inflammation.  
 XX Homo sapiens.  
 OS WO200170958-A1.  
 PN 27-SEP-2001.  
 PD 19-MAR-2001; 2001WO-CN00356.  
 XX 22-MAR-2000; 2000CN-0115043.  
 PR (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.  
 PA

Tue May 13 12:12:42 2003

CC susceptibility to these conditions and for monitoring treatment.  
 CC Precancer is also indicated by detecting post-translational  
 CC modification of the epithelial peptide which is a marker of epithelial  
 CC cell transformation. Antibodies are potentially useful for diagnosis  
 CC and treatment of cancer.

XX Sequence 16 AA;  
 SQ Query Match 7.2%; Score 7; DB 19; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 2.8; 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0

Qy 26 ERKKRE 32  
 Db 10 ERKKRE 16

RESULT 9  
 AAW50918  
 ID AAW50918 standard; peptide; 16 AA.

XX AC AAW50918;

XX DT 07-SEP-1998 (first entry)

XX DE Human immune/haematopoietic antigen SEQ ID NO:10928.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytototoxic; gene therapy; vaccine; metastasis.

XX OS Homo sapiens.

XX PN WO200157182-A2.

XX PD 09-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01354.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226681.

XX PR 22-AUG-2000; 2000US-0226868.

XX PR 22-AUG-2000; 2000US-0227182.

XX PR 23-AUG-2000; 2000US-0227009.

PI Mao Y, Xie Y;

DR WPI; 2001-602782/68.

XX New human DNA mismatch repair protein 15 for diagnosing and treating  
 PT malignant tumor, hemopathy, human immunodeficiency virus infection,  
 PT immunological diseases and inflammations

PS Example 5; Page 14; 34pp; Chinese.

XX The present invention describes human DNA mismatch repair protein 15 (I).  
 CC (I) has cytostatic, virucidal, immunomodulatory, antiinflammatory,  
 CC haemostatic and anti-HIV activities. The polynucleotide (II) encoding  
 CC (I) can be used in gene therapy. (I) and (II) can be used in the  
 CC diagnosis and treatment of malignant tumour, haemopathy, human  
 CC immunodeficiency virus (HIV) infection, immunological diseases and  
 CC various inflammations. The present sequence represents the human DNA  
 CC mismatch repair protein 15 N-terminal peptide, which is used in an  
 CC example from the present invention.

XX Sequence 15 AA;

Query Match 7.2%; Score 7; DB 22; Length 15;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 RKKERK 29

Db 9 RKKERK 15

RESULT 8

AAW50918

ID AAW50918 standard; peptide; 16 AA.

XX AC AAW50918;

XX DT 09-SEP-1998 (first entry)

XX DE Amino acid sequence of a human epithelial peptide.

XX KW Human epithelial peptide; marker; cancer; probe; hybridisation;  
 KW primer; amplification; lung; liver; kidney; breast; prostate;  
 KW melanoma; myeloma; antibody.

XX OS Homo sapiens.

XX PN WO9814469-A2.

XX PD 09-APR-1998.

XX PF 02-OCT-1997; 97WO-US17714.

XX PR 02-OCT-1996; 96US-0725027.

XX PR (UJJO ) UNIV JOHNS HOPKINS.

XX PR (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX PI Mulshine JL, Tockman MS;

XX DR WPI; 1998-240016/21.

XX New isolated epithelial protein as early marker of cancer - useful  
 PT in computer-assisted methods of diagnosis based on discriminant  
 PT analysis of optical images of cells

PS Claim 2; Page 10; 159; English.

XX This is the amino acid sequence of the human epithelial peptide, used  
 CC in the method of the invention as early markers for cancer. Probes  
 CC and primers that hybridise to or amplify these peptides are used to  
 CC diagnose precancerous states, e.g. of lung, liver, kidney, breast,  
 CC prostate, head or neck, melanoma or myeloma, or to determine

PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 06-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 08-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231143.  
 PR 08-SEP-2000; 2000US-0231144.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0232081.  
 PR 14-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0232403.  
 PR 21-SEP-2000; 2000US-0233064.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 25-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234987.  
 PR 26-SEP-2000; 2000US-0234988.  
 PR 27-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 29-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 02-OCT-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0241826.  
 PR 08-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 06-DEC-2000; 2000US-0256719.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2000US-0254097.  
 PR 05-JAN-2001; 2000US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

N-PSDB; AAK56116.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Claim 11; SEQ ID NO 10928; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/hematopoietic-related diseases, especially cancers and cancer metastases of hematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/hematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention.

Sequence 28 AA;

Query Match

Best Local Similarity 7.2%; Score 7; DB 22; Length 28;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30

Db 12 KKERKK 18

RESULT 10  
AA005715



PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234273.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 01-DEC-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 05-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 06-DEC-2000; 2000US-0256719.  
 PR 08-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.

PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-465567/50.  
 DR N-PSDB; AAS39462.  
 XX Isolated polypeptide for treating, preventing and/or prognosing  
 PT disorders related to the colon including colon cancers and also for  
 PT testing and detection e.g. diagnosis -  
 XX Claim 11; SEQ ID No 359; 562pp; English.  
 PS The present invention relates to the isolation of novel human colon  
 CC associated polypeptides, and the cDNA (AAS39348-AAS39581) and genomic  
 CC sequences encoding for them. The sequences of the invention are useful  
 CC in the diagnosis, treatment, prevention and/or prognosis of disorders  
 CC of the colon including colon cancer, congenital abnormalities  
 CC (e.g. atresia and stenosis), bacterial and viral infections,  
 CC inflammatory bowel disease (IBD), neoplastic cell disorders,  
 CC (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,  
 CC colonic inflammation, diarrhoea and dysentery, malabsorption syndromes,  
 CC (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.  
 CC The polynucleotide sequences of the invention can also be used in gene  
 CC therapy. AAU22468-AAU22701 represent the novel human colon associated  
 CC polypeptides of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Seq Sequence 36 AA;  
 SQ  
 Query Match 7.2%; Score 7; DB 22; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 6.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 24 KKERKKK 30  
 Db |||||  
 7 KKERKKK 13  
 RESULT 12  
 AAM92555  
 ID AAM92555 standard; Protein; 36 AA.  
 AC AAM92555;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX Human digestive system antigen SEQ ID NO: 1904.  
 DE Human; digestive system antigen; gene therapy; cancer; appendicitis;  
 KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
 KW digestive system disorder; Meckel's diverticulum.  
 XX Homo sapiens.  
 OS  
 XX WO200155314-A2.  
 FN  
 XX 02-AUG-2001.  
 PD  
 XX 17-JAN-2001; 2001WO-US01324.  
 PF  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.





CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer, Meckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a digestive system antigen of  
 CC the invention.  
 XX  
 SQ Sequence 36 AA;

Query Match 7.2%; Score 7; DB 22; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 6.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 KKERKK 30  
 Db 7 KKERKK 13  
 |||||

RESULT 13  
 AAO00857  
 ID AAO00857 standard; Protein; 40 AA.  
 AC AAO00857;  
 DT 06-NOV-2001 (first entry)  
 DE Human polypeptide SEQ ID NO 14749.  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 OS Homo sapiens.

XX WO200164835-A2.  
 XX 07-SEP-2001.  
 XX 26-FEB-2001; 2001WO-US04927.  
 XX 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 PA (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-514838/56.  
 DR N-PSDB; RAI80786.  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 XX diagnosing and treating e.g. leukaemia, inflammation and immune  
 XX disorders.

XX Claim 20; SEQ ID NO 14749; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 40 AA;

Query Match 7.2%; Score 7; DB 22; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 6.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 SETLSQT 20  
 Db 22 SETLSQT 28  
 |||||

RESULT 14  
 AAO08040  
 ID AAO08040 standard; Protein; 41 AA.  
 XX AAO08040;  
 AC AAO08040;  
 DT 06-NOV-2001 (first entry)  
 DE Human polypeptide SEQ ID NO 21932.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 OS Homo sapiens.

XX WO200164835-A2.  
 XX 07-SEP-2001.  
 XX 26-FEB-2001; 2001WO-US04927.  
 XX 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-514838/56.  
 DR N-PSDB; AAI87971.

XX Isolated nucleic acids and polypeptides, useful for preventing  
 XX diagnosing and treating e.g. leukaemia, inflammation and immune  
 XX disorders.

XX Claim 20; SEQ ID NO 21932; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

XX Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 41 AA;

Query Match 7.2%; Score 7; DB 22; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 RKERKK 29  
 Db 16 RKERKK 22  
 |||||

DT	XX	17-DEC-2001	(first entry)
DE	XX	Human cardiovascular system antigen polypeptide SEQ ID No 1018.	
KW	XX	Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;	
KW	XX	chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;	
KW	XX	antirheumatic; antiproliferative; cytostatic; cardiac; neuroprotective;	
KW	XX	cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;	
KW	XX	ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasia;	
KW	XX	hyperproliferative disorder; breast; liver; cardiovascular disorder;	
KW	XX	cerebrovascular disorder; nervous system disorder; bacterial infection;	
KW	XX	fungal infection; viral infection; ocular disorder; endocrine disorder;	
KW	XX	gastrointestinal disorder; renal disorder; respiratory disorder;	
KW	XX	wound healing; skin aging; organ transplantation; tissue regeneration;	
OS	XX	anti-infertility.	
PN	XX	Homo sapiens.	
PD	XX	WO200155321-A2.	
PF	XX	02-AUG-2001.	
XX	XX	17-JAN-2001; 2001WO-US01340.	
XX	XX	31-JAN-2000; 2000US-0179065.	
PR	XX	04-FEB-2000; 2000US-0180628.	
PR	XX	24-FEB-2000; 2000US-0184664.	
PR	XX	02-MAR-2000; 2000US-0186350.	
PR	XX	16-MAR-2000; 2000US-0189874.	
PR	XX	17-MAR-2000; 2000US-0190076.	
PR	XX	18-APR-2000; 2000US-0198123.	
PR	XX	19-MAY-2000; 2000US-0205515.	
PR	XX	07-JUN-2000; 2000US-0209467.	
PR	XX	28-JUN-2000; 2000US-0214886.	
PR	XX	30-JUN-2000; 2000US-0215135.	
PR	XX	07-JUL-2000; 2000US-0216647.	
PR	XX	07-JUL-2000; 2000US-0216880.	
PR	XX	11-JUL-2000; 2000US-0217487.	
PR	XX	14-JUL-2000; 2000US-0217496.	
PR	XX	26-JUL-2000; 2000US-0218290.	
PR	XX	26-JUL-2000; 2000US-0220964.	
PR	XX	14-AUG-2000; 2000US-0224518.	
PR	XX	14-AUG-2000; 2000US-0224519.	
PR	XX	14-AUG-2000; 2000US-0225213.	
PR	XX	14-AUG-2000; 2000US-0225214.	
PR	XX	14-AUG-2000; 2000US-0225266.	
PR	XX	14-AUG-2000; 2000US-0225267.	
PR	XX	14-AUG-2000; 2000US-0225268.	
PR	XX	14-AUG-2000; 2000US-0225270.	
PR	XX	14-AUG-2000; 2000US-0225447.	
PR	XX	14-AUG-2000; 2000US-0225757.	
PR	XX	14-AUG-2000; 2000US-0225758.	
PR	XX	14-AUG-2000; 2000US-0225759.	
PR	XX	18-AUG-2000; 2000US-0226279.	
PR	XX	22-AUG-2000; 2000US-0226681.	
PR	XX	22-AUG-2000; 2000US-0226868.	
PR	XX	22-AUG-2000; 2000US-0227182.	
PR	XX	23-AUG-2000; 2000US-0227009.	
PR	XX	30-AUG-2000; 2000US-0228924.	
PR	XX	01-SEP-2000; 2000US-0229287.	
PR	XX	01-SEP-2000; 2000US-0229343.	
PR	XX	01-SEP-2000; 2000US-0229344.	
PR	XX	01-SEP-2000; 2000US-0229345.	
PR	XX	05-SEP-2000; 2000US-0229509.	
PR	XX	05-SEP-2000; 2000US-0229513.	
PR	XX	06-SEP-2000; 2000US-0230437.	
PR	XX	06-SEP-2000; 2000US-0230438.	
PR	XX	08-SEP-2000; 2000US-0231242.	
PR	XX	08-SEP-2000; 2000US-0231243.	
PR	XX	08-SEP-2000; 2000US-0231244.	
PR	XX	08-SEP-2000; 2000US-0231413.	
PR	XX	08-SEP-2000; 2000US-0231414.	
PR	XX	08-SEP-2000; 2000US-0231415.	
RESULT 15			
ID	AAU08046	standard; Protein; 43 AA.	
XX	AAU08046		
XX	AAU08046		
XX	06-NOV-2001	(first entry)	
DE	XX	Human polypeptide SEQ ID NO 21938.	
XX	XX	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	XX	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	
KW	XX	tissue growth factor; immunomodulatory; cancer; leukaemia;	
KW	XX	nervous system disorders; arthritis; inflammation.	
OS	XX	Homo sapiens.	
PN	XX	WO200164835-A2.	
XX	XX	07-SEP-2001.	
XX	XX	26-FEB-2001; 2001WO-US04927.	
XX	XX	28-FEB-2000; 2000US-0515126.	
PR	XX	18-MAY-2000; 2000US-0577409.	
XX	XX	(HYSE-) HYSEQ INC.	
PA	XX	Tang YT, Liu C, Drmanac RT;	
PI	XX	WPI; 2001-514838/56.	
XX	XX	N-PSDB; AAI87977.	
DR	XX	Isolated nucleic acids and polypeptides, useful for preventing	
PT	XX	diagnosing and treating e.g. leukaemia, inflammation and immune	
PT	XX	disorders.	
PT	XX	Claim 20; SEQ ID NO 21938; 1399pp + Sequence Listing; English.	
XX	XX	The invention relates to human polynucleotides (AAI79941-AAI93841) and	
CC	XX	the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to	
CC	XX	cytokine, cell proliferation or cell differentiation or which may induce	
CC	XX	production of other cytokines in other cell populations. The	
CC	XX	polynucleotides and polypeptides are useful in gene therapy, vaccines or	
CC	XX	peptide therapy. The polypeptides have various cytokine-like activities,	
CC	XX	e.g. stem cell growth factor activity, haematopoiesis regulating	
CC	XX	activity, tissue growth factor activity, immunomodulatory activity and	
CC	XX	activin/inhibin activity and may be useful in the diagnosis and/or	
CC	XX	treatment of cancer, leukaemia, nervous system disorders, arthritis and	
CC	XX	inflammation.	
CC	XX	Note: The sequence data for this patent did not form part of the printed	
CC	XX	specification, but was obtained in electronic format directly from WIPO	
CC	XX	at ftp.wipo.int/pub/published_pct_sequences.	
XX			

PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0232403.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234224.  
 PR 25-SEP-2000; 2000US-0234277.  
 PR 25-SEP-2000; 2000US-0234297.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0234984.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 01-NOV-2000; 2000US-0241826.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249219.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 03-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2000US-0254097.  
 XX 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI; 2001-451930/48.  
 DR N-PSDB; AAS35518.  
 XX

New cardiovascular system related polynucleotides and polypeptides,  
 useful for diagnosing, treating and/or preventing disorders of the  
 cardiovascular system

Claim 11; SEQ ID No 1018; 674pp; English.

Sequences AAU21852-AAU22466 represent the cardiovascular system antigen  
 polypeptides of the invention. Cardiovascular system antigens and their  
 associated polynucleotides are useful in the diagnosis, treatment and  
 prevention of various types of disorders in e.g. humans, mice, rabbits,  
 goats, horses, cats, dogs, chickens or sheep. A pathological condition  
 can be determined by detecting the presence or absence of a mutation in a  
 cardiovascular system antigen polynucleotide. The treatable disorders  
 include autoimmune diseases such as rheumatoid arthritis,  
 hyperproliferative disorders such as neoplasms of the breast or liver,  
 cardiovascular disorders such as cerebral ischaemia, nervous system disorders  
 such as Alzheimer's disease, infections caused by bacteria, viruses and fungi,  
 ocular disorders such as corneal infection, endocrine disorders such as  
 premature labour and infertility, gastrointestinal disorders such as  
 Crohn's disease, renal disorders such as glomerulonephritis and  
 respiratory disorders such as asthma and pleurisy. The polypeptides can  
 also be used to aid wound healing, to prevent skin aging due to sunburn,  
 to maintain organs before transplantation, to regenerate tissues and in  
 chemotaxis.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match

Best Local Similarity 7.2%; Score 7; DB 22; Length 47;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31

Db 14 KERKKR 20

RESULT 17

AAG02455

ID AAG02455 standard; Protein; 52 AA.

XX AAG02455;

XX 06-OCT-2000 (first entry)

Human secreted protein, SEQ ID NO: 6536.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 gene therapy; chromosome mapping.

Tue May 13 12:12:42 2003

OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 XX 21-FEB-2000; 2000EP-0200610.  
 PF  
 XX 26-FEB-1999; 99US-0122487.  
 PR  
 XX (GEST ) GENSET.  
 XX  
 XX Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI WPI; 2000-500381/45.  
 DR N-PSDB; AAC02461.  
 DR  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PT  
 XX Claim 13; SEQ ID 6536; 71pp + CD-ROM; English.  
 PS  
 XX The present sequence is a polypeptide encoded by one of a large number  
 XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 XX  
 XX Sequence 52 AA;  
 SQ  
 Query Match 7.2%; Score 7; DB 21; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKK 29  
 DB 18 RKKERKK 24  
 RESULT 18  
 AAG00160  
 ID AAG00160 standard; Protein; 54 AA.  
 XX  
 AC AAG00160;  
 XX  
 XX 06-OCT-2000 (first entry)  
 DT  
 XX Human secreted protein, SEQ ID NO: 4241.  
 DE  
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 KW  
 XX Homo sapiens.  
 OS  
 XX EP1033401-A2.  
 PN  
 XX 06-SEP-2000.  
 PD  
 XX 21-FEB-2000; 2000EP-0200610.  
 PF  
 XX 26-FEB-1999; 99US-0122487.  
 PR  
 XX (GEST ) GENSET.  
 PA  
 XX

PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX WPI; 2000-500381/45.  
 DR N-PSDB; AAC00166.  
 XX  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PT  
 XX Claim 13; SEQ ID 4241; 71pp + CD-ROM; English.  
 PS  
 XX The present sequence is a polypeptide encoded by one of a large number  
 XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 XX  
 XX Sequence 54 AA;  
 SQ  
 Query Match 7.2%; Score 7; DB 21; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 8.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 SLGDSSET 16  
 DB 45 SLGDSSET 51  
 RESULT 19  
 AAO09451  
 ID AAO09451 standard; Protein; 57 AA.  
 XX  
 AC AAO09451;  
 XX  
 XX 06-NOV-2001 (first entry)  
 DT  
 XX Human polypeptide SEQ ID NO 23343.  
 DE  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 OS  
 XX WO200164835-A2.  
 PN  
 XX 07-SEP-2001.  
 PD  
 XX 26-FEB-2001; 2001WO-US04927.  
 PF  
 XX 28-FEB-2000; 2000US-0515126.  
 PR  
 XX 18-MAY-2000; 2000US-0577409.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-514838/56.  
 PI N-PSDB; AAI89382.  
 DR  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 PT  
 XX

PS Claim 20; SEQ ID NO 23343; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 57 AA;

Query Match 7.2%; Score 7; DB 22; Length 57;  
 Best Local Similarity 100.0%; Pred. No. 9.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29  
 Db 4 RKKERKK 10

RESULT 20  
 AAO10486  
 ID AAO10486 standard; Protein; 59 AA.

AC AAO10486;

XX  
 DT 06-NOV-2001 (first entry)  
 XX Human polypeptide SEQ ID NO 24378.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX

OS Homo sapiens.

XX WO200164835-A2.

PN 07-SEP-2001.

PD 26-FEB-2001; 2001WO-US04927.

PF 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI90417.

PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders.

XX Claim 20; SEQ ID NO 24378; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 59 AA;

Query Match 7.2%; Score 7; DB 22; Length 59;  
 Best Local Similarity 100.0%; Pred. No. 9.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29  
 Db 46 RKKERKK 52

RESULT 21  
 AAO01774  
 ID AAO01774 standard; Protein; 60 AA.

XX AAO01774;

XX 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 15666.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX

OS Homo sapiens.

XX WO200164835-A2.

PN 07-SEP-2001.

PD 26-FEB-2001; 2001WO-US04927.

PF 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI81705.

PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders.

XX Claim 20; SEQ ID NO 15666; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

us-09-854-133-586.olig.rag

Tue May 13 12:12:42 2003

XX Sequence 60 AA;  
SQ

Query Match 7.2%; Score 7; DB 22; Length 60;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29  
Db 35 RKKERKK 41

RESULT 22  
AAO02325  
ID AAO02325 standard; Protein; 62 AA.

XX AAO02325;  
AC AAO02325;

XX 06-NOV-2001 (first entry)  
DT

XX Human polypeptide SEQ ID NO 16217.  
DE

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.  
OS

XX WO200164835-A2.  
PN

XX 07-SEP-2001.  
PD

XX 26-FEB-2001; 2001WO-US04927.  
PF

XX 28-FEB-2000; 2000US-0515126.  
PR

XX 18-MAY-2000; 2000US-0577409.  
PR

XX (HYSE-) HYSEQ INC.  
PA

XX Tang YT, Liu C, Drmanac RT;  
PI

XX WPI; 2001-514838/56.  
DR

XX N-PSDB; AAI82256.  
DR

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

XX Claim 20; SEQ ID NO 16217; 1399pp + Sequence Listing; English.  
PS

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 62 AA;  
SQ

Query Match 7.2%; Score 7; DB 22; Length 62;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29

|||||||  
Db 50 RKKERKK 56

RESULT 23

AAO09233  
ID AAO09233 standard; Protein; 64 AA.

XX AAO09233;  
AC AAO09233;

XX 06-NOV-2001 (first entry)  
DT

XX Human polypeptide SEQ ID NO 23125.  
DE

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.  
OS

XX WO200164835-A2.  
PN

XX 07-SEP-2001.  
PD

XX 26-FEB-2001; 2001WO-US04927.  
PF

XX 28-FEB-2000; 2000US-0515126.  
PR

XX 18-MAY-2000; 2000US-0577409.  
PR

XX (HYSE-) HYSEQ INC.  
PA

XX Tang YT, Liu C, Drmanac RT;  
PI

XX WPI; 2001-514838/56.  
DR

XX N-PSDB; AAI89164.  
DR

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

XX Claim 20; SEQ ID NO 23125; 1399pp + Sequence Listing; English.  
PS

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 64 AA;  
SQ

Query Match 7.2%; Score 7; DB 22; Length 64;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29

Db 16 RKKERKK 22

RESULT 24

ABB03142  
ID ABB03142 standard; Protein; 66 AA.

XX

AC ABB03142;  
XX  
DT 08-JAN-2002 (first entry)  
XX  
DE Human musculoskeletal system related polypeptide SEQ ID NO 1089.  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system.  
XX  
OS Homo sapiens.  
PN WO200155367-A1.  
XX  
XX 02-AUG-2001.  
PD  
XX  
PF 17-JAN-2001; 2001WO-US01338.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 25-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 02-OCT-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.





XX PD 07-SEP-2001.  
 XX PF 26-FEB-2001; 2001WO-US04927.  
 XX PR 28-FEB-2000; 2000US-0515126.  
 XX PR 18-MAY-2000; 2000US-0577409.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX DR N-PSDB; AAI85266.  
 XX DR WPI; 2001-514938/56.  
 XX DR N-PSDB; AAI85266.  
 XX PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders.  
 XX PS Claim 20; SEQ ID NO 19227; 1399pp + Sequence Listing; English.  
 XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX CC  
 XX SQ Sequence 80 AA;  
 Query Match 7.2%; Score 7; DB 22; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKK 29  
 Db 52 RKKERKK 58  
 |||||  
 RESULT 27  
 ABB04590  
 ID ABB04590 standard; Protein; 81 AA.  
 AC ABB04590;  
 XX AC  
 XX DT 21-MAR-2002 (first entry)  
 XX DE Human peroxidase 9.  
 XX KW Human; peroxidase 9; cancer; haemopathy; HIV infection; phlogosis;  
 KW cytostatic; haemostatic; virucide; immunomodulatory; antiinflammatory;  
 KW immune disease; gene therapy; protein; enzyme.  
 XX OS Homo sapiens.  
 XX PA WO200192317-A1.  
 XX PN 06-DEC-2001.  
 XX PD 21-MAY-2001; 2001WO-CN00844.  
 XX PF 24-MAY-2000; 2000CN-0115848.  
 XX PR (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.  
 XX PA  
 XX CC

PI Mao Y, Xie Y;  
 XX WPI; 2002-090029/12.  
 DR N-PSDB; ABA05891.  
 XX DR  
 XX PT HOMO peroxidase 9 and encoding polynucleotide, used in diagnosis and  
 PT treatment of malignant tumors, hemopathy, human immunodeficiency virus  
 PT infection, immunological diseases and inflammation -  
 XX Claim 1; Page 29; 37pp; Chinese.  
 XX CC The present invention provides the protein and coding sequences of human  
 CC peroxidase 9. The sequences can be used in the treatment of cancer,  
 CC haemopathy, HIV infection, phlogosis and immune diseases. The present  
 CC sequence is the protein of the invention.  
 XX CC  
 XX SQ Sequence 81 AA;  
 Query Match 7.2%; Score 7; DB 23; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKK 29  
 Db 21 RKKERKK 27  
 |||||  
 RESULT 28  
 ABG10474  
 ID ABG10474 standard; Protein; 82 AA.  
 XX AC  
 XX AC ABG10474;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #10465.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX DR WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS74661.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX PS Claim 20; SEQ ID No 40833; 103pp; English.  
 XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or

Tue May 13 12:12:42 2003

quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 7.2%; Score 7; DB 22; Length 82;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKERKK 29  
Db 17 RKERKK 23

RESULT 29  
ABP10332  
ID ABP10332 standard; Protein; 82 AA.  
XX  
AC ABP10332;  
DT 25-JUN-2002 (first entry)  
XX Human ORFX protein sequence SEQ ID NO:20646.  
DE  
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX Homo sapiens.  
OS  
XX WO200192523-A2.  
PN  
XX 06-DEC-2001.  
PD  
XX 29-MAY-2001; 2001WO-US10836.  
PF  
XX 30-MAY-2000; 2000US-206132P.  
PR  
XX 29-AUG-2000; 2000US-228716P.  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX Shimkets RA, Leach MD;  
PI  
XX WPI; 2002-106308/14.  
DR  
XX N-PSDB; ABN26084.  
DR  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders  
PT  
XX Disclosure; SEQ ID 20646; 1037pp; English.  
PS  
XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated

disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative diseases, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 7.2%; Score 7; DB 23; Length 82;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30  
Db 18 KKERKK 24

RESULT 30  
ABB72352  
ID ABB72352 standard; Protein; 94 AA.  
XX  
AC ABB72352;  
XX  
DT 04-APR-2002 (first entry)  
DE Murine protein isolated from skin cells SEQ ID NO: 676.  
XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;  
KW developmental defect; inflammatory disease; dermatological; vulvar;  
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective.  
XX Mus sp.  
OS  
XX WO200190357-A1.  
PN  
XX 29-NOV-2001.  
PD  
XX 24-MAY-2001; 2001WO-NZ00099.  
PF  
XX 24-MAY-2000; 2000US-206650P.  
PR  
XX 25-JUL-2000; 2000US-221232P.  
XX  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA  
XX Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD;  
PI  
XX WPI; 2002-122020/16.  
DR  
XX N-PSDB; ABL35041.  
DR  
XX New polynucleotides and polypeptides encoded by the polynucleotides  
PT isolated from skin cells, useful for treating skin wounds, cancers,  
PT growth and developmental defects, inflammatory diseases, or for  
PT modulating immune responses  
PT  
XX Claim 4; Page 432; 466pp; English.  
PS  
XX The present invention provides the protein and coding sequences of CDNAs  
CC isolated from human, murine and rat skin cell libraries. The sequences

CC can be used in the development of therapeutic agents useful in the  
 CC treatment of skin diseases, including skin wounds, cancer, growth  
 CC defects, developmental defects and inflammatory diseases. The proteins  
 CC have important roles in the induction of hair growth, cell proliferation  
 CC and cell-cell interaction, in maintaining tissue integrity, in wound  
 CC healing and in modulating immune responses. The present sequence is a  
 CC polypeptide of the invention.

SQ Sequence 94 AA;

Query Match 7.2%; Score 7; DB 23; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKRRKK 29  
 Db 35 RKRRKK 41  
 |||||

RESULT 31  
 AAO03592  
 ID AAO03592 standard; Protein; 96 AA.

AC AAO03592;  
 DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 17484.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 tissue growth factor; immunomodulatory; cancer; leukaemia;  
 nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;  
 WPI; 2001-514838/56.

DR N-PSDB; AAI83523.

Isolated nucleic acids and polypeptides, useful for preventing  
 diagnosing and treating e.g. leukaemia, inflammation and immune  
 disorders -

Claim 20; SEQ ID NO 17484; 1399pp + Sequence Listing; English.

The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 cytokine, cell proliferation or cell differentiation or which may induce  
 production of other cytokines in other cell populations. The  
 polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 peptide therapy. The polypeptides have various cytokine-like activities,  
 e.g. stem cell growth factor activity, haematopoiesis regulating  
 activity, tissue growth factor activity, immunomodulatory activity and  
 activin/inhibin activity and may be useful in the diagnosis and/or  
 treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 inflammation.

Note: The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 96 AA;

Query Match 7.2%; Score 7; DB 22; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKRRKK 29  
 Db 24 RKRRKK 30  
 |||||

RESULT 32  
 AAO09213  
 ID AAO09213 standard; Protein; 102 AA.

AC AAO09213;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 23105.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 tissue growth factor; immunomodulatory; cancer; leukaemia;  
 nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;  
 WPI; 2001-514838/56.

DR N-PSDB; AAI89144.

Isolated nucleic acids and polypeptides, useful for preventing  
 diagnosing and treating e.g. leukaemia, inflammation and immune  
 disorders -

Claim 20; SEQ ID NO 23105; 1399pp + Sequence Listing; English.

The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 cytokine, cell proliferation or cell differentiation or which may induce  
 production of other cytokines in other cell populations. The  
 polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 peptide therapy. The polypeptides have various cytokine-like activities,  
 e.g. stem cell growth factor activity, haematopoiesis regulating  
 activity, tissue growth factor activity, immunomodulatory activity and  
 activin/inhibin activity and may be useful in the diagnosis and/or  
 treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 inflammation.

Note: The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 102 AA;

Query Match 7.2%; Score 7; DB 22; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKRRKK 29  
 |||||

us-09-854-133-586.olig.rag

Tue May 13 12:12:42 2003

```

Db      29 RKKKK 35

RESULT 33
AA010463
ID      AA010463 standard; Protein; 102 AA.
XX
XX      AA010463;
AC
XX      06-NOV-2001 (first entry)
DT
XX
XX      Human polypeptide SEQ ID NO 24355.
DE
XX      Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW      vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW      tissue growth factor; immunomodulatory; cancer; leukaemia;
KW      nervous system disorders; arthritis; inflammation.
XX
XX      Homo sapiens.
OS
XX      WO200164835-A2.
PN
XX      07-SEP-2001.
PD
XX      26-FEB-2001; 2001WO-US04927.
PF
XX      28-FEB-2000; 2000US-0515126.
XX      18-MAY-2000; 2000US-0577409.
XX
XX      (HYSE-) HYSEQ INC.
PA
XX      Tang YT, Liu C, Drmanac RT;
PI
XX      WPI; 2001-514838/56.
XX      N-PSDB; AA190394.
DR
XX      Isolated nucleic acids and polypeptides, useful for preventing
PT      diagnosing and treating e.g. leukaemia, inflammation and immune
PT      disorders.
XX
XX      Claim 20; SEQ ID NO 24355; 1399pp + Sequence Listing; English.
PS
XX      The invention relates to human polynucleotides (AA179941-AA193841) and
XX      the encoded proteins (AA000010-AA013910) that exhibit activity elating to
XX      cytokine, cell proliferation or cell differentiation or which may induce
XX      production of other cytokines in other cell populations. The
XX      polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX      peptide therapy. The polypeptides have various cytokine-like activities,
XX      e.g. stem cell growth factor activity, haematopoiesis regulating
XX      activity, tissue growth factor activity, immunomodulatory activity and
XX      activin/inhibin activity and may be useful in the diagnosis and/or
XX      treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX      inflammation.
XX      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Query Match      7.2%; Score 7; DB 22; Length 102;
XX      Best Local Similarity 100.0%; Pred. No. 16;
XX      Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY      24 RKKKK 30
XX      Db      57 RKKKK 63
XX
XX      RESULT 34
XX      AAG12424
XX      ID      AAG12424 standard; Protein; 104 AA.
XX
XX      AAG12424;
AC

```

```

XX      17-OCT-2000 (first entry)
XX
XX      Zea mays protein fragment SEQ ID NO: 11530.
DE
XX
XX      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence; corn.
XX
XX      Zea mays subsp. mays.
OS
XX      EP1033405-A2.
PN
XX      06-SEP-2000.
PD
XX
XX      25-FEB-2000; 2000EP-0301439.
PF
XX      25-FEB-1999; 99US-0121825.
XX      05-MAR-1999; 99US-0123180.
XX      09-MAR-1999; 99US-0123548.
XX      23-MAR-1999; 99US-0125788.
XX      25-MAR-1999; 99US-0126264.
XX      29-MAR-1999; 99US-0126785.
XX      01-APR-1999; 99US-0127462.
XX      06-APR-1999; 99US-0128234.
XX      08-APR-1999; 99US-0128714.
XX      16-APR-1999; 99US-0129845.
XX      19-APR-1999; 99US-0130077.
XX      21-APR-1999; 99US-0130449.
XX      23-APR-1999; 99US-0130510.
XX      28-APR-1999; 99US-0130891.
XX      30-APR-1999; 99US-0131449.
XX      30-APR-1999; 99US-0132048.
XX      04-MAY-1999; 99US-0132407.
XX      05-MAY-1999; 99US-0132484.
XX      06-MAY-1999; 99US-0132485.
XX      06-MAY-1999; 99US-0132486.
XX      07-MAY-1999; 99US-0132487.
XX      11-MAY-1999; 99US-0132863.
XX      14-MAY-1999; 99US-0134256.
XX      14-MAY-1999; 99US-0134218.
XX      14-MAY-1999; 99US-0134219.
XX      14-MAY-1999; 99US-0134221.
XX      14-MAY-1999; 99US-0134270.
XX      18-MAY-1999; 99US-0134768.
XX      19-MAY-1999; 99US-0134941.
XX      20-MAY-1999; 99US-0135124.
XX      21-MAY-1999; 99US-0135353.
XX      24-MAY-1999; 99US-0135629.
XX      25-MAY-1999; 99US-0136021.
XX      27-MAY-1999; 99US-0136392.
XX      28-MAY-1999; 99US-0136782.
XX      01-JUN-1999; 99US-0137222.
XX      03-JUN-1999; 99US-0137528.
XX      04-JUN-1999; 99US-0137502.
XX      07-JUN-1999; 99US-0137724.
XX      08-JUN-1999; 99US-0138094.
XX      10-JUN-1999; 99US-0138540.
XX      10-JUN-1999; 99US-0138847.
XX      14-JUN-1999; 99US-0139119.
XX      16-JUN-1999; 99US-0139452.
XX      16-JUN-1999; 99US-0139453.
XX      17-JUN-1999; 99US-0139492.
XX      18-JUN-1999; 99US-0139454.
XX      18-JUN-1999; 99US-0139455.
XX      18-JUN-1999; 99US-0139456.
XX      18-JUN-1999; 99US-0139457.
XX      18-JUN-1999; 99US-0139458.
XX      18-JUN-1999; 99US-0139459.
XX      18-JUN-1999; 99US-0139460.
XX      18-JUN-1999; 99US-0139461.
XX      18-JUN-1999; 99US-0139462.
XX      18-JUN-1999; 99US-0139463.
XX

```

```

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

```

```

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 7.28; Score 7; DB 21; Length 104;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 24 KKERKKK 30
Db 37 KKERKKK 43

```

```

RESULT 35
ABG10473
ID ABG10473 standard; Protein; 109 AA.
XX
AC ABG10473;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #10464.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.

```

us-09-854-133-586.olig.rag

Tue May 13 12:12:42 2003

```

XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS90694.
XX XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity
XX PS Claim 20; SEQ ID No 56866; 103pp; English.
XX PS The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 109 AA;

Query Match 7.2%; Score 7; DB 22; Length 109;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 RKKERKK 29
Db 82 RKKERKK 88
|||||

RESULT 37
AAC02869
ID AAC02869 standard; Protein; 111 AA.
XX AC
XX AC AAC02869;
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 16761.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.

```

XX PN WO200164835-A2.  
 XX PD 07-SEP-2001.  
 XX PF 26-FEB-2001; 2001WO-US04927.  
 XX PR 28-FEB-2000; 2000US-0515126.  
 XX PR 18-MAY-2000; 2000US-0577409.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX PR WPI; 2001-514838/56.  
 XX DR N-PSDB; AAI82800.  
 XX PT Isolated nucleic acids and polypeptides, useful for preventing  
 XX PT disorders -  
 XX PS Claim 20; SEQ ID NO 16761; 1399pp + Sequence Listing; English.  
 XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity, and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 111 AA;  
 Query Match 7.2%; Score 7; DB 22; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERK 29  
 Db 19 RKKERK 25  
 RESULT 38  
 AAE01505  
 ID AAE01505 standard; Protein; 111 AA.  
 AC AAE01505;  
 XX 17-JUL-2001 (first entry)  
 DT Human gene 22 encoded secreted protein HE9SE18, SEQ ID NO:162.  
 DE Human: secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnery;  
 KW cell culture; chemotaxis; food additive; gene therapy;  
 KW binding partner identification.  
 XX OS Homo sapiens.  
 XX

FH Key Location/Qualifiers  
 FT Peptide 1  
 FT Protein /label= Signal\_peptide  
 FT 2..111  
 XX /label= Human\_mature\_secreted\_protein  
 PN WO200134626-A1.  
 XX PD 17-MAY-2001.  
 XX PF 01-NOV-2000; 2000WO-US30045.  
 XX PR 05-NOV-1999; 99US-0163581.  
 XX PR 30-JUN-2000; 2000US-0215133.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Ruben SM, Komatsoulis GA, Moore PA, Birse CE, N1 J;  
 XX DR WPI; 2001-308778/32.  
 XX DR N-PSDB; AAD05371.  
 XX PT New nucleic acid molecules encoding 28 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 PS Claim 11; Page 527-528; 562pp; English.  
 XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
 CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
 CC AAE01514-AAE01544 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the  
 CC amount of the new protein in a sample or by determining the presence of  
 CC mutations in the new genes. Specific uses are described for each of the  
 CC 28 genes, based on the tissues in which they are most highly expressed,  
 CC and include developing products for the diagnosis or treatment of  
 CC proliferative disorders, cancer, tumours, foetal and developmental  
 CC abnormalities, haematopoietic disorders, diseases of the immune system,  
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
 CC disorders, and infections. The proteins can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues, to identify their  
 CC cognate ligands or binding partners, and in chemotaxis, and can be used  
 CC as a food additive or preservative to modify storage properties.  
 CC Antibodies specific for a protein of the invention can be used in  
 CC alleviating symptoms associated with the disorders mentioned above, and  
 CC in diagnostic immunoassays e.g. radioimmunoassay or enzyme linked  
 CC immunosorbent assay (ELISA). The present sequence represents a human  
 CC secreted protein of the invention.  
 XX SQ Sequence 111 AA;  
 Query Match 7.2%; Score 7; DB 22; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERK 29  
 Db 92 RKKERK 98  
 RESULT 39  
 ABG63908  
 ID ABG63908 standard; Protein; 111 AA.  
 XX



Tue May 13 12:12:42 2003

Human polypeptide SEQ ID NO 18478.

DE Human polypeptide SEQ ID NO 18478.  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX Homo sapiens.  
 XX WO200164835-A2.  
 PN 07-SEP-2001.  
 PD 26-FEB-2001; 2001WO-US04927.  
 PF 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Dermanac RT;  
 PI WPI; 2001-514838/56.  
 DR N-PSDB; AAI84517.  
 DR Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX Claim 20; SEQ ID NO 18478; 1399pp + Sequence Listing; English.  
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and/or  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 114 AA;  
 Query Match 7.2%; Score 7; DB 22; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 24 KKRRKK 30  
 Db 81 KKRRKK 87  
 RESULT 41  
 AAO01695  
 ID AAO01695 standard; Protein; 118 AA.  
 XX AC AAO01695;  
 XX 06-NOV-2001 (first entry)  
 XX Human polypeptide SEQ ID NO 15587.  
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX Homo sapiens.  
 OS

AC ABG63908;  
 XX 27-AUG-2002 (first entry)  
 DE Human albumin fusion protein #583.  
 XX Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KW human serum albumin; HSA; cancer; reproductive disorder;  
 KW digestive disorder; immune disorder; endocrine disorder;  
 KW haematopoietic disorder; neural disorder; connective disorder;  
 KW cytostatic; anti-infectivity; anti-inflammatory; antitumor;  
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;  
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritic.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX WO200177137-A1.  
 PN 18-OCT-2001.  
 PD 12-APR-2001; 2001WO-US11988.  
 PF 12-APR-2000; 2000US-229358P.  
 PR 25-APR-2000; 2000US-199384P.  
 PR 21-DEC-2000; 2000US-256931P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Haseltine WA;  
 PI WPI; 2002-010886/01.  
 DR New fusion protein for treating disease e.g. diabetes comprises an  
 XX albumin fused to a therapeutic protein -  
 XX Claim 1; Page 899-900; 2102pp; English.  
 XX The present invention relates to albumin fusion proteins comprising a  
 CC therapeutic protein X and human albumin (HA, also known as human serum  
 CC albumin, HSA). The proteins are useful for treating a disease or  
 CC disorder that may be modulated by therapeutic protein X. The albumin  
 CC extends the shelf-life of protein X, and may increase its biological  
 CC in vitro/in vivo activity. The protein is useful for treating and  
 CC diagnosing disorders such as cancer, reproductive disorders, digestive  
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders  
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders  
 CC (e.g. diabetes), haematopoietic disorders, neural disorders  
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,  
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders  
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin  
 CC fusion proteins of the invention.  
 XX SQ Sequence 111 AA;  
 Query Match 7.2%; Score 7; DB 23; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKRRKK 29  
 Db 92 RKRRKK 98  
 RESULT 40  
 AAO04586  
 ID AAO04586 standard; Protein; 114 AA.  
 XX AC AAO04586;  
 XX 06-NOV-2001 (first entry)  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX Homo sapiens.  
 OS

XX WO200164835-A2.  
 XX 07-SEP-2001.  
 XX 26-FEB-2001; 2001WO-US04927.  
 XX 28-FEB-2000; 2000US-0515126.  
 XX 18-MAY-2000; 2000US-0577409.  
 XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-514838/56.  
 XX N-PSDB; AAI81626.  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 XX disorders -  
 XX Claim 20; SEQ ID NO 15587; 1399pp + Sequence Listing; English.  
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 XX cytokine, cell proliferation or cell differentiation or which may induce  
 XX production of other cytokines in other cell populations. The  
 XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 XX peptide therapy. The polypeptides have various cytokine-like activities,  
 XX e.g. stem cell growth factor activity, haematopoiesis regulating  
 XX activity, tissue growth factor activity, immunomodulatory regulating  
 XX activity/inhibin activity and may be useful in the diagnosis and/or  
 XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 XX inflammation.  
 XX Note: The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 118 AA;

Query Match 7.2%; Score 7; DB 22; Length 118;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KKKRERK 34  
 Db 31 KKKRERK 37

RESULT 42  
 AAG27487  
 ID AAG27487 standard; Protein; 120 AA.  
 AC AAG27487;  
 XX 17-OCT-2000 (first entry)  
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 32343.  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 XX hybridisation assay; genetic mapping; gene expression control; promoter;  
 XX termination sequence.  
 XX Arabidopsis thaliana.  
 XX EP1033405-A2.  
 XX 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 XX 25-FEB-1999; 99US-0121825.  
 XX 05-MAR-1999; 99US-0123180.

us-09-854-133-586.olig.rag

Tue May 13 12:12:42 2003

```

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.2%; Score 7; DB 21; Length 120;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KKKRERK 34
DB 96 KKKRERK 102
|||||

RESULT 43
ABP43004
ID ABP43004 standard; Protein; 121 AA.
XX AC ABP43004;
XX DT 22-AUG-2002 (first entry)
XX DE Human ovarian antigen HSAYW59, SEQ ID NO:4136.
XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
ovarian cancer; breast cancer; tumour; reproductive system disorder;
infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
inflammatory condition; immune disorder; blood disorder;
cardiovascular disorder; respiratory disorder; neurological disorder;
gastrointestinal disorder; urinary system disorder; drug screening;
gene therapy; chromosome mapping; forensic analysis;
antibody preparation; cytostatic; immunomodulatory; neuroprotective;
antiinflammatory; gynaecological; reproductive.
XX OS Homo sapiens.
XX PN WO200200677-A1.
XX PD 03-JAN-2002.
XX PF 07-JUN-2001; 2001WO-US18569.
XX PR 07-JUN-2000; 2000US-209467P.
XX XX

```

PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Birse CE, Rosen CA;  
 XX WPI; 2002-147878/19.  
 DR N-PSDB; ABQ56081.  
 XX  
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
 PT ovarian cancer), immune disorders, cardiovascular disorders and  
 PT neurological diseases -  
 XX  
 PS Claim 11; SEQ ID No 4136; 2922pp; English.  
 XX  
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigen  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovarian cancer and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents a human ovarian antigen of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 121 AA;

Query Match 7.2%; Score 7; DB 23; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKK 29  
 Db 26 RKKERKK 32  
 |||||

RESULT 44  
 ABG07519  
 ID ABG07519 standard; Protein; 124 AA.  
 XX  
 AC ABG07519;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #7510.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.  
 XX  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Llu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS71706.  
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 37878; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 124 AA;

Query Match 7.2%; Score 7; DB 22; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKK 29  
 Db 31 RKKERKK 37  
 |||||

RESULT 45  
 AAG54869  
 ID AAG54869 standard; Protein; 132 AA.  
 XX  
 AC AAG54869;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 70139.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX

us-09-854-133-586.olig.rag

Tue May 13 12:12:42 2003

PF	25-FEB-2000; 2000EP-0301439.	09-JUL-1999;	99US-0142920.
XX		PR	99US-0142977.
PR	25-FEB-1999;	PR	99US-0143542.
PR	05-MAR-1999;	PR	99US-0144005.
PR	09-MAR-1999;	PR	99US-0144085.
PR	23-MAR-1999;	PR	99US-0144086.
PR	09-MAR-1999;	PR	99US-0144325.
PR	29-MAR-1999;	PR	99US-0144331.
PR	01-APR-1999;	PR	99US-0144332.
PR	06-APR-1999;	PR	99US-0144333.
PR	16-APR-1999;	PR	99US-0144335.
PR	21-APR-1999;	PR	99US-0144352.
PR	23-APR-1999;	PR	99US-0144632.
PR	28-APR-1999;	PR	99US-0144884.
PR	30-APR-1999;	PR	99US-0144814.
PR	04-MAY-1999;	PR	99US-0145086.
PR	06-MAY-1999;	PR	99US-0145088.
PR	07-MAY-1999;	PR	99US-0145085.
PR	11-MAY-1999;	PR	99US-0145087.
PR	14-MAY-1999;	PR	99US-0145089.
PR	14-MAY-1999;	PR	99US-0145192.
PR	14-MAY-1999;	PR	99US-0145145.
PR	14-MAY-1999;	PR	99US-0145218.
PR	14-MAY-1999;	PR	99US-0145224.
PR	18-MAY-1999;	PR	99US-0145276.
PR	19-MAY-1999;	PR	99US-0145913.
PR	21-MAY-1999;	PR	99US-0145918.
PR	21-MAY-1999;	PR	99US-0145919.
PR	25-MAY-1999;	PR	99US-0145951.
PR	25-MAY-1999;	PR	99US-0145951.
PR	25-MAY-1999;	PR	99US-0146386.
PR	25-MAY-1999;	PR	99US-0146388.
PR	25-MAY-1999;	PR	99US-0146389.
PR	25-MAY-1999;	PR	99US-0147038.
PR	25-MAY-1999;	PR	99US-0147038.
PR	25-MAY-1999;	PR	99US-0147204.
PR	25-MAY-1999;	PR	99US-0147302.
PR	25-MAY-1999;	PR	99US-0147192.
PR	25-MAY-1999;	PR	99US-0147260.
PR	25-MAY-1999;	PR	99US-0147303.
PR	25-MAY-1999;	PR	99US-0147416.
PR	25-MAY-1999;	PR	99US-0147935.
PR	25-MAY-1999;	PR	99US-0147935.
PR	25-MAY-1999;	PR	99US-0148171.
PR	25-MAY-1999;	PR	99US-0148319.
PR	25-MAY-1999;	PR	99US-0148341.
PR	25-MAY-1999;	PR	99US-0148565.
PR	25-MAY-1999;	PR	99US-0148684.
PR	25-MAY-1999;	PR	99US-0149368.
PR	25-MAY-1999;	PR	99US-0149175.
PR	25-MAY-1999;	PR	99US-0149426.
PR	25-MAY-1999;	PR	99US-0149722.
PR	25-MAY-1999;	PR	99US-0149723.
PR	25-MAY-1999;	PR	99US-0149929.
PR	25-MAY-1999;	PR	99US-0149902.
PR	25-MAY-1999;	PR	99US-0149930.
PR	25-MAY-1999;	PR	99US-0150566.
PR	25-MAY-1999;	PR	99US-0150884.
PR	25-MAY-1999;	PR	99US-0151065.
PR	25-MAY-1999;	PR	99US-0151066.
PR	25-MAY-1999;	PR	99US-0151080.
PR	25-MAY-1999;	PR	99US-0151303.
PR	25-MAY-1999;	PR	99US-0151303.
PR	25-MAY-1999;	PR	99US-0151338.
PR	25-MAY-1999;	PR	99US-0151930.
PR	25-MAY-1999;	PR	99US-0152363.
PR	25-MAY-1999;	PR	99US-0153070.
PR	25-MAY-1999;	PR	99US-0153758.
PR	25-MAY-1999;	PR	99US-0154018.
PR	25-MAY-1999;	PR	99US-0154039.
PR	25-MAY-1999;	PR	99US-0154779.
PR	25-MAY-1999;	PR	99US-0155139.
PR	25-MAY-1999;	PR	99US-0155486.
PR	25-MAY-1999;	PR	99US-0155659.
PR	25-MAY-1999;	PR	99US-0155659.

PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match

Best Local Similarity 7.2%; Score 7; DB 21; Length 132;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KKKRERK 34  
Db 107 KKKRERK 113

Search completed: May 11, 2003, 20:10:48  
Job time : 58.6549 secs

us-09-854-133-586.olig.ra1

Tue May 13 12:12:42 2003

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:09:32 ; Search time 24.0354 Seconds  
(without alignments)  
118.743 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 97  
Sequence: 1 EVEVSRDHASLGDSFTLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0 262574

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database :  
1: /cgn2.6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2.6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2.6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2.6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2.6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	6.2	7	4	US-09-258-754-407
2	6	6.2	7	4	US-09-042-107-407
3	6	6.2	27	6	5171680-14
4	6	6.2	73	4	US-09-250-609-19
5	6	6.2	88	4	US-09-134-001C-2855
6	6	6.2	113	1	US-08-168-091A-26
7	6	6.2	163	4	US-09-562-737-77
8	6	6.2	184	2	US-08-715-204-5
9	6	6.2	184	2	US-08-691-814B-50
10	6	6.2	184	3	US-09-162-597-5
11	6	6.2	184	4	US-09-250-609-13
12	6	6.2	185	2	US-08-691-814B-12
13	6	6.2	185	4	US-09-250-609-11
14	6	6.2	246	3	US-08-675-885-2
15	6	6.2	252	4	US-09-134-001C-3847
16	6	6.2	319	1	US-08-702-344-28
17	6	6.2	331	2	US-08-646-981-17
18	6	6.2	348	4	US-08-853-948B-5
19	6	6.2	350	4	US-09-651-200-25
20	6	6.2	399	1	US-08-414-926A-5
21	6	6.2	399	2	US-08-928-922-5
22	6	6.2	399	3	US-09-253-682-5
23	6	6.2	399	4	US-09-527-657-5
24	6	6.2	439	4	US-09-134-001C-4903
25	6	6.2	503	4	US-09-562-737-70
26	6	6.2	540	2	US-08-724-394A-4
27	6	6.2	635	4	US-08-931-608A-5

28	6	6.2	641	4	US-09-071-035-456	Sequence 456, App
29	6	6.2	775	2	US-08-714-070A-1	Sequence 1, Appl
30	6	6.2	1313	4	US-09-071-035-450	Sequence 450, App
31	6	6.2	1313	4	US-09-071-035-454	Sequence 454, App
32	6	6.2	1447	4	US-09-041-886-25	Sequence 25, Appl
33	6	6.2	1447	5	PCT-US94-05277-2	Sequence 2, Appl
34	6	6.2	2629	2	US-08-751-189-4	Sequence 4, Appl
35	6	6.2	2629	2	US-09-060-836-4	Sequence 4, Appl
36	6	6.2	2629	2	US-09-184-445-4	Sequence 3, Appl
37	6	6.2	3816	1	US-09-428-517-3	Sequence 2, Appl
38	5	5.2	6	4	US-08-439-747A-2	Sequence 105, App
39	5	5.2	6	4	US-08-440-409B-2	Sequence 1, Appl
40	5	5.2	6	4	US-09-025-596-105	Sequence 1, Appl
41	5	5.2	8	2	US-08-701-124-1	Sequence 1, Appl
42	5	5.2	8	2	US-09-130-225-1	Sequence 74, Appl
43	5	5.2	8	4	US-09-455-061-1	Sequence 94, Appl
44	5	5.2	8	4	US-09-101-751A-74	Sequence 23, Appl
45	5	5.2	8	4	US-09-101-751A-94	Sequence 23, Appl
46	5	5.2	9	1	US-08-439-747A-23	Sequence 23, Appl
47	5	5.2	9	2	US-08-440-409B-23	Sequence 281, App
48	5	5.2	9	3	US-08-159-339A-281	Sequence 291, App
49	5	5.2	10	3	US-08-159-339A-291	Sequence 19, Appl
50	5	5.2	11	2	US-08-701-124-19	Sequence 6, Appl
51	5	5.2	11	2	US-08-623-690-6	Sequence 19, Appl
52	5	5.2	11	3	US-09-130-225-19	Sequence 19, Appl
53	5	5.2	11	4	US-09-455-061-19	Sequence 76, Appl
54	5	5.2	11	4	US-09-101-751A-76	Sequence 76, Appl
55	5	5.2	11	6	5200183-7	Patent No. 5200183
56	5	5.2	11	6	5200183-10	Patent No. 5200183
57	5	5.2	12	1	US-08-477-509B-16	Sequence 16, Appl
58	5	5.2	12	2	US-08-623-690-7	Sequence 7, Appl
59	5	5.2	12	3	US-08-623-690-16	Sequence 16, Appl
60	5	5.2	12	4	US-08-482-085B-16	Sequence 16, Appl
61	5	5.2	13	1	US-09-444-791A-16	Sequence 30, Appl
62	5	5.2	13	1	US-08-787-547-30	Sequence 8, Appl
63	5	5.2	13	6	US-08-623-690-8	Patent No. 5168050
64	5	5.2	15	1	US-07-720-189-7	Sequence 7, Appl
65	5	5.2	15	1	US-07-720-189-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-09-258-754-407  
; Sequence 407, Application US/09258754  
; Patent No. 6174687  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; APPLICANT: Rajotte, Daniel  
; TITLE OF INVENTION: Methods of Identifying Lung Homing Molecules Using  
; TITLE OF INVENTION: Membrane Dipeptidase  
; FILE REFERENCE: P-LJ 3443  
; CURRENT APPLICATION NUMBER: US/09/258,754  
; EARLIER FILING DATE: 1999-02-26  
; EARLIER FILING DATE: 1998-03-13  
; NUMBER OF SEQ ID NOS: 452  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 407  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-258-754-407

Query Match 6.2%; Score 6; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 78 LTGGCL 83

Db 2 LTGGCL 7  
|||||

## RESULT 2

US-09-042-107-407  
; Sequence 407, Application US/09042107  
; Patent No. 6232287  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; TITLE OF INVENTION: Molecules that Home to Various Selected Organs or  
; FILE REFERENCE: P-LJ 2892  
; CURRENT FILING DATE: 1998-03-13  
; CURRENT APPLICATION NUMBER: US/09/042,107  
; NUMBER OF SEQ ID NOS: 436  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 407  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-042-107-407

## Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 7;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 LTGGCL 83

Db 2 LTGGCL 7  
|||||

## RESULT 3

US-09-042-107-407  
; Patent No. 5171680  
; APPLICANT: MULLENBACH, GUY T.; HALLEWELL, ROBERT A.; VALEZUELA,  
; PABLO  
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE ANALOGS HAVING NOVEL  
; BINDING PROPERTIES  
; NUMBER OF SEQUENCES: 15  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/561,442  
; FILING DATE: 01-AUG-1990  
; SEQ ID NO:14:  
; LENGTH: 27  
5171680-14

## Query Match

Best Local Similarity 6.2%; Score 6; DB 6; Length 27;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKR 31

Db 13 ERKKR 18  
|||||

## RESULT 4

US-09-250-609-19  
; Sequence 19, Application US/09250609A  
; Patent No. 6458943  
; GENERAL INFORMATION:  
; APPLICANT: Byrnie, Jennifer A.  
; TITLE OF INVENTION: Members of the D52 Gene Family  
; FILE REFERENCE: 1383.0210002  
; CURRENT APPLICATION NUMBER: US/09/250,609A  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 73

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-250-609-19

## Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 73;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19

Db 16 SETLSQ 21  
|||||

## RESULT 5

US-09-134-001C-2855  
; Sequence 2855, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2855  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2855

## Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 88;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERK 28

Db 43 RKKERK 48  
|||||

## RESULT 6

US-08-168-091A-26  
; Sequence 26, Application US/08168091A  
; Patent No. 5665862  
; GENERAL INFORMATION:  
; APPLICANT: Fischbach, Gerald.  
; APPLICANT: Falls, Douglas R.  
; APPLICANT: Rosen, Kenneth M.  
; APPLICANT: Corfas, Gabriel  
; TITLE OF INVENTION: Neurotrophic Factor  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE AND COCKFIELD  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/168,091A  
; FILING DATE: 15-DEC-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/953,742



us-09-854-133-586.olig.ra1

Tue May 13 12:12:42 2003

FILING DATE: 29-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: DeConti, Giulio A  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: HMI-002CP  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 113 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-168-091A-26

Query Match 6.2%; Score 6; DB 1; Length 113;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 29  
DB 92 KKERKK 97

RESULT 7  
US-09-562-737-77  
Sequence 77, Application US/09562737  
Patent No. 642867  
GENERAL INFORMATION:  
APPLICANT: Herz, Joachim  
APPLICANT: Gotthardt, Michael  
TITLE OF INVENTION: LDL Receptor Signaling Pathways  
FILE REFERENCE: UTSW0708  
CURRENT APPLICATION NUMBER: US/09/562,737  
CURRENT FILING DATE: 2000-05-01  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 77  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-562-737-77

Query Match 6.2%; Score 6; DB 4; Length 163;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGDS 14  
DB 13 ASLGDS 18

RESULT 8  
US-08-715-204-5  
Sequence 5, Application US/08715204  
Patent No. 5874286  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Au-Young, Janice  
APPLICANT: Goli, Surya K.  
APPLICANT: Hillman, Jennifer.  
APPLICANT: Zweiger, Gary B.  
TITLE OF INVENTION: A NOVEL TUMOR PROTEIN  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/715,204  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0126 US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 790225  
US-08-715-204-5

Query Match 6.2%; Score 6; DB 2; Length 184;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
DB 100 SETLSQ 105

RESULT 9  
US-08-691-814B-50  
Sequence 50, Application US/08691814B  
Patent No. 5981218  
GENERAL INFORMATION:  
APPLICANT: Rio, Marie-Christine  
APPLICANT: Tomasetto, Catherine  
APPLICANT: Basset, Paul  
APPLICANT: Byrne, Jennifer  
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful  
TITLE OF INVENTION: as Leukemia Markers and in Breast Cancer Prognosis  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Ave, NW, Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/691,814B  
FILING DATE: 31-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/002,183

;; FILING DATE: 09-AUG-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Steffe, Eric K.  
;; REGISTRATION NUMBER: 36,688  
;; REFERENCE/DOCKET NUMBER: 1383.00900001  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-371-2600  
;; TELEFAX: 202-371-2543  
;; INFORMATION FOR SEQ ID NO: 50:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 184 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: not relevant  
;; TOPOLOGY: not relevant  
;; MOLECULE TYPE: peptide  
US-08-691-814B-50

Query Match 6.2%; Score 6; DB 2; Length 184;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
Db 100 SETLSQ 105

## RESULT 10

US-09-162-597-5  
; Sequence 5, Application US/09162597  
; Patent No. 6043343

## GENERAL INFORMATION:

;; APPLICANT: Bandman, Olga  
;; APPLICANT: Au-Young, Janice  
;; APPLICANT: Goli, Surya K.  
;; APPLICANT: Hillman, Jennifer.  
;; APPLICANT: Zweigler, Gary B.  
;; TITLE OF INVENTION: A NOVEL TUMOR PROTEIN  
;; NUMBER OF SEQUENCES: 7  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
;; STREET: 3174 Porter Drive  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: U.S.  
;; ZIP: 94304

## COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq Version 1.5  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/162,597  
;; FILING DATE:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/715,204  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Billings, Lucy J.  
;; REGISTRATION NUMBER: 36,749  
;; REFERENCE/DOCKET NUMBER: PF-0126 US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-855-0555  
;; TELEFAX: 415-845-4166  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 184 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; IMMEDIATE SOURCE:  
;; LIBRARY: GenBank  
;; CLONE: 790225

## US-09-162-597-5

Query Match 6.2%; Score 6; DB 3; Length 184;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
Db 100 SETLSQ 105

## RESULT 11

US-09-250-609-13  
; Sequence 13, Application US/09250609A  
; Patent No. 6458943

## GENERAL INFORMATION:

;; APPLICANT: Byrne, Jennifer A.  
;; TITLE OF INVENTION: Members of the D52 Gene Family  
;; FILE REFERENCE: 1383.0210002  
;; CURRENT APPLICATION NUMBER: US/09/250,609A  
;; CURRENT FILING DATE: 1999-02-17  
;; NUMBER OF SEQ ID NOS: 108  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 13  
;; LENGTH: 184  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-250-609-13

## Query Match

6.2%; Score 6; DB 4; Length 184;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
Db 100 SETLSQ 105

## RESULT 12

US-08-691-814B-12  
; Sequence 12, Application US/08691814B  
; Patent No. 5981218

## GENERAL INFORMATION:

;; APPLICANT: Rio, Marie-Christine  
;; APPLICANT: Tomasetto, Catherine  
;; APPLICANT: Basset, Paul  
;; APPLICANT: Byrne, Jennifer  
;; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful  
;; TITLE OF INVENTION: as Leukemia Markers and in Breast Cancer Prognosis  
;; NUMBER OF SEQUENCES: 124  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
;; STREET: 1100 New York Ave, NW, Suite 600  
;; CITY: Washington  
;; STATE: DC  
;; COUNTRY: USA  
;; ZIP: 20005-3934

## COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/691,814B  
;; FILING DATE: 31-JUL-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/002,183  
;; FILING DATE: 09-AUG-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Steffe, Eric K.  
;; REGISTRATION NUMBER: 36,688  
;; REFERENCE/DOCKET NUMBER: 1383.00900001

us-09-854-133-586.olig.ra1

Tue May 13 12:12:42 2003

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-371-2600  
 TELEFAX: 202-371-2543  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 185 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-691-814B-12

Query Match 6.2%; Score 6; DB 2; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 SETLSQ 19  
 Db 100 SETLSQ 105

RESULT 13  
 US-09-250-609-11  
 ; Sequence 11, Application US/09250609A  
 ; Patent No. 6458943  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Byrne, Jennifer A.  
 ; TITLE OF INVENTION: Members of the D52 Gene Family  
 ; FILE REFERENCE: 1383.0210002  
 ; CURRENT APPLICATION NUMBER: US/09/250,609A  
 ; CURRENT FILING DATE: 1999-02-17  
 ; NUMBER OF SEQ ID NOS: 108  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 11  
 ; LENGTH: 185  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-250-609-11

Query Match 6.2%; Score 6; DB 4; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 SETLSQ 19  
 Db 100 SETLSQ 105

RESULT 14  
 US-08-675-885-2  
 ; Sequence 2, Application US/08675885  
 ; Patent No. 6066723  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Grammatikakis, Nicholas  
 ; APPLICANT: Grammatikakis, Aliki  
 ; APPLICANT: Toole, Bryan P.  
 ; APPLICANT: Cochran, Brent  
 ; TITLE OF INVENTION: NUCLEIC ACID ENCODING VERTEBRATE CDC37  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kevin M. Farrell, P.C.  
 ; STREET: P.O. Box 999  
 ; CITY: York Harbor  
 ; STATE: ME  
 ; COUNTRY: US  
 ; ZIP: 03909  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/675,885  
 ; FILING DATE:

CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fairrell, Kevin M.  
 REGISTRATION NUMBER: 35,505  
 REFERENCE/DOCKET NUMBER: TU-9601  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 207 3630558  
 TELEFAX: 207 3630528  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 246 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-675-885-2

Query Match 6.2%; Score 6; DB 3; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 ELRKE 26  
 Db 3 ELRKE 8

RESULT 15  
 US-09-134-001C-3847  
 ; Sequence 3847, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 3847  
 ; LENGTH: 252  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-3847

Query Match 6.2%; Score 6; DB 4; Length 252;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 KKKRR 33  
 Db 25 KKKRR 30

RESULT 16  
 US-08-702-344-28  
 ; Sequence 28, Application US/08702344  
 ; Patent No. 5723315  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jacobs, Kenneth  
 ; APPLICANT: McCoy, John  
 ; APPLICANT: Lavallie, Edward  
 ; APPLICANT: Racie, Lisa  
 ; APPLICANT: Merberg, David  
 ; APPLICANT: Treacy, Maurice  
 ; APPLICANT: Spaulding, Vikki  
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
 ; TITLE OF INVENTION: ENCODING THEM  
 ; NUMBER OF SEQUENCES: 37  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/702,344  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 319 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-702-344-28

Query Match 6.2%; Score 6; DB 1; Length 319;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTE 21  
DB 247 TLSQTE 252

RESULT 17  
US-08-646-981-17  
; Sequence 17, Application US/08646981  
; Patent No. 5852183  
; GENERAL INFORMATION:  
; APPLICANT: MAEDA, HIROAKI  
; APPLICANT: ED, YASUYUKI  
; APPLICANT: KIMACHI, KAZUHIKO  
; APPLICANT: ONO, YOICHI  
; APPLICANT: TOKIYOSHI, SACHIO  
; TITLE OF INVENTION: DOG-MOUSE HETEROHYBRIDOMA AND GENE  
; TITLE OF INVENTION: FRAGMENT CODING FOR CONSTANT REGION OF CANINE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: BIRCH, STEWART, KOLASCH AND BIRCH  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/646,981  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WEINER, MARC S  
; REGISTRATION NUMBER: 32,181  
; REFERENCE/DOCKET NUMBER: 1488-106  
; INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:  
LENGTH: 331 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-646-981-17

Query Match 6.2%; Score 6; DB 2; Length 331;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPK 71  
DB 233 PPSPK 238

RESULT 18  
US-08-853-948B-5  
; Sequence 5, Application US/08853948B  
; Patent No. 6210943  
; GENERAL INFORMATION:  
; APPLICANT: AKIHAMA, Toyota  
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING  
; TITLE OF INVENTION: THE SAME  
; FILE REFERENCE: 0049-0235-0  
; CURRENT APPLICATION NUMBER: US/08/853,948B  
; CURRENT FILING DATE: 1997-05-09  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Citrus unshiu  
US-08-853-948B-5

Query Match 6.2%; Score 6; DB 4; Length 348;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 EMLTGG 81  
DB 42 EMLTGG 47

RESULT 19  
US-09-651-200-25  
; Sequence 25, Application US/09651200  
; Patent No. 6429303  
; GENERAL INFORMATION:  
; APPLICANT: Green et al  
; TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B  
; TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and  
; TITLE OF INVENTION: Polypeptides Encoded Thereby  
; FILE REFERENCE: 15966-562 (CURA-62)  
; CURRENT APPLICATION NUMBER: US/09/651,200  
; CURRENT FILING DATE: 2000-08-30  
; PRIOR APPLICATION NUMBER: 60/152383  
; PRIOR FILING DATE: 1999-09-03  
; PRIOR APPLICATION NUMBER: 60/172909  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/183578  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 350  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-651-200-25

Query Match 6.2%; Score 6; DB 4; Length 350;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKRE 32

|||||

Db 280 RKKRE 285

# RESULT 20

US-08-414-926A-5

; Sequence 5, Application US/08414926A

; Patent No. 5721354

; GENERAL INFORMATION:

; APPLICANT: Spaete, Richard

; APPLICANT: Cha, Tai-An

; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum

; STREET: 5 Palo Alto Square

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94306-2155

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/414,926A

; FILING DATE: March 31, 1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Cseri, Luann

; REGISTRATION NUMBER: 31,822

; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-494-7622

; TELEFAX: 415-857-0663

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 399 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-414-926A-5

Query Match 6.2%; Score 6; DB 1; Length 399;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 FIIFWI 48

|||||

Db 20 FIIFWI 25

# RESULT 21

US-08-926-922-5

; Sequence 5, Application US/08926922

; Patent No. 5925751

; GENERAL INFORMATION:

; APPLICANT: Spaete, Richard

; APPLICANT: Cha, Tai-An

; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Luann Cseri Attorney at Law

; STREET: 750 Arimo Avenue

; CITY: Oakland

; STATE: CA

; COUNTRY: USA

; ZIP: 94610

; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/926,922

FILING DATE: September 10, 1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR 11A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-834-1448

TELEFAX: 510-839-7810

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 399 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-926-922-5

Query Match 6.2%; Score 6; DB 2; Length 399;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 FIIFWI 48

|||||

Db 20 FIIFWI 25

# RESULT 22

US-09-253-682-5

; Sequence 5, Application US/09253682

; Patent No. 6040170

; GENERAL INFORMATION:

; APPLICANT: Spaete, Richard

; APPLICANT: Cha, Tai-An

; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Luann Cseri Attorney at Law

; STREET: 750 Arimo Avenue

; CITY: Oakland

; STATE: CA

; COUNTRY: USA

; ZIP: 94610

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/253,682

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/926,922

FILING DATE: September 10, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR 11A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-834-1448

TELEFAX: 510-839-7810

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 399 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein.

US-09-253-682-5

Query Match  
Best Local Similarity 6.2%; Score 6; DB 3; Length 399;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 FIIFWI 48  
Db 20 FIIFWI 25

RESULT 23

US-09-527-657-5  
Sequence 5, Application US/09527657  
Patent No. 6291236

GENERAL INFORMATION:

APPLICANT: Spaete, Richard

Cha, Tai-An

TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Luann Cserr Attorney at Law

STREET: 750 Arimo Avenue

CITY: Oakland

STATE: CA

COUNTRY: USA

ZIP: 94610

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/527,657

FILING DATE: 17-Mar-2000

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/926,922

FILING DATE: September 10, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR 11A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-834-1448

TELEFAX: 510-839-7810

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 399 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-527-657-5

Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 399;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 FIIFWI 48  
Db 20 FIIFWI 25

RESULT 24

US-09-134-001C-4903

Sequence 4903, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

STREET: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4903  
LENGTH: 439  
TYPE: PRT

ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4903

Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 439;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKK 25  
Db 31 TELRKK 36

RESULT 25

US-09-562-737-70

Sequence 70, Application US/09562737

Patent No. 6428967

GENERAL INFORMATION:

APPLICANT: Herz, Joachim

APPLICANT: Gotthardt, Michael

TITLE OF INVENTION: LDL Receptor Signaling Pathways

FILE REFERENCE: UTSW0708

CURRENT APPLICATION NUMBER: US/09/562,737

CURRENT FILING DATE: 2000-05-01

NUMBER OF SEQ ID NOS: 132

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 70

LENGTH: 503

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic

OTHER INFORMATION: Sequence

US-09-562-737-70

Query Match

Best Local Similarity 6.2%; Score 6; DB 4; Length 503;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 LGRKK 96  
Db 416 LGRKK 421

RESULT 26

US-08-724-394A-4

Sequence 4, Application US/08724394A

Patent No. 5872337

GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta

APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA

us-09-854-133-586.olig.rai

Tue May 13 12:12:42 2003

```

;
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELEPHONE: 415-576-0300
; TELEFAX: 415-576-0200
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 540 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..540
; OTHER INFORMATION: /note= "BTF5"
;
; US-08-724-394A-4
;
; Query Match 6.2%; Score 6; DB 2; Length 540;
; Best Local Similarity 100.0%; Pred. No. 2e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 27 RKKRE 32
;      |||||
; Db 298 RKKRE 303
;
; RESULT 27
; US-08-931-608A-5
; Sequence 5, Application US/08931608A
; Patent No. 6302685
; GENERAL INFORMATION:
; APPLICANT: Lobel, Peter
; APPLICANT: Sleat, David E.
; TITLE OF INVENTION: NOVEL HUMAN LYSOSOMAL PROTEIN AND METHODS OF ITS USE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,608A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 601-1-077
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 5:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
;
; US-08-931-608A-5
;
; Query Match 6.2%; Score 6; DB 4; Length 635;
; Best Local Similarity 100.0%; Pred. No. 2.3e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 9 ASLGS 14
;      |||||
; Db 194 ASLGS 199
;
; RESULT 28
; US-09-071-035-456
; Sequence 456, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 456:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 641 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-09-071-035-456
;
; Query Match 6.2%; Score 6; DB 4; Length 641;
; Best Local Similarity 100.0%; Pred. No. 2.3e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 15 ETLST 20
;      |||||
; Db 241 ETLST 246
;
; RESULT 29
; US-08-714-070A-1
; Sequence 1, Application US/08714070A
```

```
; Patent No. 5834237
; GENERAL INFORMATION:
; APPLICANT: JACOBS, Eric
; APPLICANT: SILVESTRE, Nathalie
; APPLICANT: SCHWEINRYBER, Ernst
; TITLE OF INVENTION: COMBINED USE OF TWO EXPRESSION CASSETTES
; TITLE OF INVENTION: FOR THE PRODUCTION OF A PROTEIN OF INTEREST
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,070A
; FILING DATE: 07-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94/01767
; FILING DATE: 10-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 017753-077
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 775 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Schizosaccharomyces pombe
; US-08-714-070A-1

Query Match 6.2%; Score 6; DB 2; Length 775;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 ERKKR 31
Db 20 ERKKR 25

RESULT 30
US-09-071-035-450
; Sequence 450, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 454:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
```

```
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 450:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-450

Query Match 6.2%; Score 6; DB 4; Length 1313;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 ETLST 20
Db 877 ETLST 882

RESULT 31
US-09-071-035-454
; Sequence 454, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 454:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
```



us-09-854-133-586.olig.rai

Tue May 13 12:12:42 2003

```
;
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-454
    Query Match          6.2%; Score 6; DB 4; Length 1313;
    Best Local Similarity 100.0%; Pred. No. 4.8e+02;
    Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ETLSTQ 20
    |||||
Db 877 ETLSTQ 882

RESULT 32
US-09-041-886-25
; Sequence 25, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharro
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-25
    Query Match          6.2%; Score 6; DB 4; Length 1447;
    Best Local Similarity 100.0%; Pred. No. 4.8e+02;
    Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGDS 14
    |||||
Db 119 ASLGDS 124

RESULT 33
PCT-US94-05277-2
; Sequence 2, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert
```

```
;
; APPLICANT: Zabrecky, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05277
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42709
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-05277-2
    Query Match          6.2%; Score 6; DB 5; Length 1447;
    Best Local Similarity 100.0%; Pred. No. 4.8e+02;
    Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGDS 14
    |||||
Db 119 ASLGDS 124

RESULT 34
US-08-751-189-4
; Sequence 4, Application US/08751189
; Patent No. 5919656
; GENERAL INFORMATION:
; APPLICANT: Harrington, Lea A.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen, Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,189
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
```

; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2629 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-751-189-4

Query Match  
Best Local Similarity 6.2%; Score 6; DB 2; Length 2629;  
Matches 100.0%; Pred. No. 8.1e+02;  
Mismatches 0; Indels 0; Gaps 0;

QY 70 KEVTCR 75  
|||||  
Db 1077 KEVTCR 1082

## RESULT 35

US-09-060-836-4  
; Sequence 4, Application US/09060836  
; Patent No. 5981707  
; GENERAL INFORMATION:

; APPLICANT: Harrington, Lea A.  
; APPLICANT: Robinson, Murray O.  
; TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/060,836  
; FILING DATE:  
; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/751,189  
; FILING DATE: 15-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.

; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2629 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-09-060-836-4

Query Match  
Best Local Similarity 6.2%; Score 6; DB 2; Length 2629;  
Matches 100.0%; Pred. No. 8.1e+02;  
Mismatches 0; Indels 0; Gaps 0;

QY 70 KEVTCR 75  
|||||  
Db 1077 KEVTCR 1082

## RESULT 36

US-09-184-445-4  
; Sequence 4, Application US/09184445

; Patent No. 6174703  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, Lea A.  
; APPLICANT: Robinson, Murray O.  
; TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/184,445  
; FILING DATE:  
; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/751,189  
; FILING DATE: 15-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2629 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-09-184-445-4

Query Match  
Best Local Similarity 6.2%; Score 6; DB 4; Length 2629;  
Matches 100.0%; Pred. No. 8.1e+02;  
Mismatches 0; Indels 0; Gaps 0;

QY 70 KEVTCR 75  
|||||  
Db 1077 KEVTCR 1082

RESULT 37  
US-09-428-517-3  
; Sequence 3, Application US/09428517  
; Patent No. 6251636  
; GENERAL INFORMATION:  
; APPLICANT: Betlach, Mary C.  
; APPLICANT: Shah, Sanjay Krishnakant  
; APPLICANT: McDaniel, Robert  
; APPLICANT: Tang, Li  
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30062-20029.00  
; CURRENT APPLICATION NUMBER: US/09/428,517  
; CURRENT FILING DATE: 1999-10-28  
; EARLIER APPLICATION NUMBER: 60/120,254  
; EARLIER FILING DATE: 1999-02-16  
; EARLIER APPLICATION NUMBER: 60/106,100  
; EARLIER FILING DATE: 1998-10-29  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 3816  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant

us-09-854-133-586.olig.ra

Tue May 13 12:12:42 2003

TITLE OF INVENTION: Soluble Active Hepatitis C Virus Protease  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Schering Corp.  
 STREET: 2000 Galloping Hill Road  
 CITY: Kenilworth  
 STATE: New Jersey  
 COUNTRY: USA  
 ZIP: 07033-0530

Query Match 6.2%; Score 6; DB 4; Length 3816;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGD 14  
 Db 1396 ASLGD 1401

RESULT 38  
 US-08-439-747A-2  
 ; Sequence 2, Application US/08439747A  
 ; Patent No. 5767233  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zhang, Rumin  
 ; APPLICANT: Murray, Michael  
 ; APPLICANT: Ramanathan, Lata  
 ; TITLE OF INVENTION: Soluble, Cleavable Substrates of the Hepatitis  
 ; C Protease  
 ; NUMBER OF SEQUENCES: 34  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Schering Corp.  
 ; STREET: 2000 Galloping Hill Road  
 ; CITY: Kenilworth  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07033-0530

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: Apple Macintosh  
 OPERATING SYSTEM: Macintosh 7.5.3  
 SOFTWARE: Microsoft Word 5.1a  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/439,747A  
 FILING DATE: May 12, 1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lunn, Paul G.  
 REGISTRATION NUMBER: 32,743  
 REFERENCE/DOCKET NUMBER: JB0509  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 908-298-5061  
 TELEFAX: 908-298-5388

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 6 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-439-747A-2

Query Match 5.2%; Score 5; DB 1; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
 Db 1 RKKR 5

RESULT 39  
 US-08-440-409B-2  
 ; Sequence 2, Application US/08440409B  
 ; Patent No. 5843752  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dasmahapatra, Bimal  
 ; APPLICANT: Murray, Michael  
 ; APPLICANT: Ramanathan, Lata  
 ; APPLICANT: Butkiewicz, Nancy

Query Match 5.2%; Score 5; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
 Db 1 RKKR 5

RESULT 40  
 US-09-025-596-105  
 ; Sequence 105, Application US/09025596  
 ; Patent No. 6340463  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mitchell, William M.  
 ; APPLICANT: Stratton, Charles W.  
 ; TITLE OF INVENTION: IDENTIFICATION OF ANTIGENIC PEPTIDE  
 ; SEQUENCE CHARACTERISTICS:  
 ; FILE REFERENCE: VDB98-01  
 ; CURRENT APPLICATION NUMBER: US/09/025,596  
 ; CURRENT FILING DATE: 1998-02-18  
 ; EARLIER APPLICATION NUMBER: 08/911,593  
 ; EARLIER FILING DATE: 1997-08-14  
 ; EARLIER APPLICATION NUMBER: 60/023,921  
 ; EARLIER FILING DATE: 1996-08-14  
 ; NUMBER OF SEQ ID NOS: 118  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 105  
 ; LENGTH: 6  
 ; TYPE: PPT  
 ; ORGANISM: Chlamydia pneumoniae  
 ; US-09-025-596-105

Query Match 5.2%; Score 5; DB 4; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKR 27  
 Db 1 RKKR 5

Db 1 RKKR 5

RESULT 41  
US-08-701-124-1  
; Sequence 1, Application US/08701124  
; Patent No. 5846782  
; GENERAL INFORMATION:  
; APPLICANT: Wickham, Thomas J.  
; APPLICANT: Roelvink, Petrus W.  
; APPLICANT: Kovesdi, Imre  
; TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza - 49th Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/087701.124  
; FILING DATE: 21-AUG-1996  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-701-124-1

Query Match 5.2%; Score 5; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
Db 1 RKKR 5

RESULT 42  
US-09-130-225-1  
; Sequence 1, Application US/09130225  
; Patent No. 6057155  
; GENERAL INFORMATION:  
; APPLICANT: Wickham, Thomas J.  
; APPLICANT: Roelvink, Petrus W.  
; APPLICANT: Kovesdi, Imre  
; TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza - 49th Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09130.225  
; FILING DATE:

QY 27 RKKR 31  
Db 1 RKKR 5

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 8-701124  
; FILING DATE: 21-AUG-1996  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-130-225-1  
Query Match 5.2%; Score 5; DB 3; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
Db 1 RKKR 5

RESULT 43  
US-09-455-061-1  
; Sequence 1, Application US/09455061  
; Patent No. 6329190  
; GENERAL INFORMATION:  
; APPLICANT: Wickham, Thomas J.  
; APPLICANT: Roelvink, Petrus W.  
; APPLICANT: Kovesdi, Imre  
; TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza - 49th Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09455.061  
; FILING DATE: 06-DEC-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 9-130225  
; FILING DATE: 06-AUG-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 8-701124  
; FILING DATE: 21-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hefner, M. Daniel  
; REGISTRATION NUMBER: 41,826  
; REFERENCE/DOCKET NUMBER: 203128  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-455-061-1

Query Match 5.2%; Score 5; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKR 31  
Db 1 RKKR 5

us-09-854-133-586.olig.ra1

Tue May 13 12:12:42 2003

OTHER INFORMATION: Description of Unknown Organism: Artificial  
OTHER INFORMATION: Sequence  
US-09-101-751A-94

Query Match 5.2%; Score 5; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKKR 31  
| | | | |  
Db 1 RKKKR 5

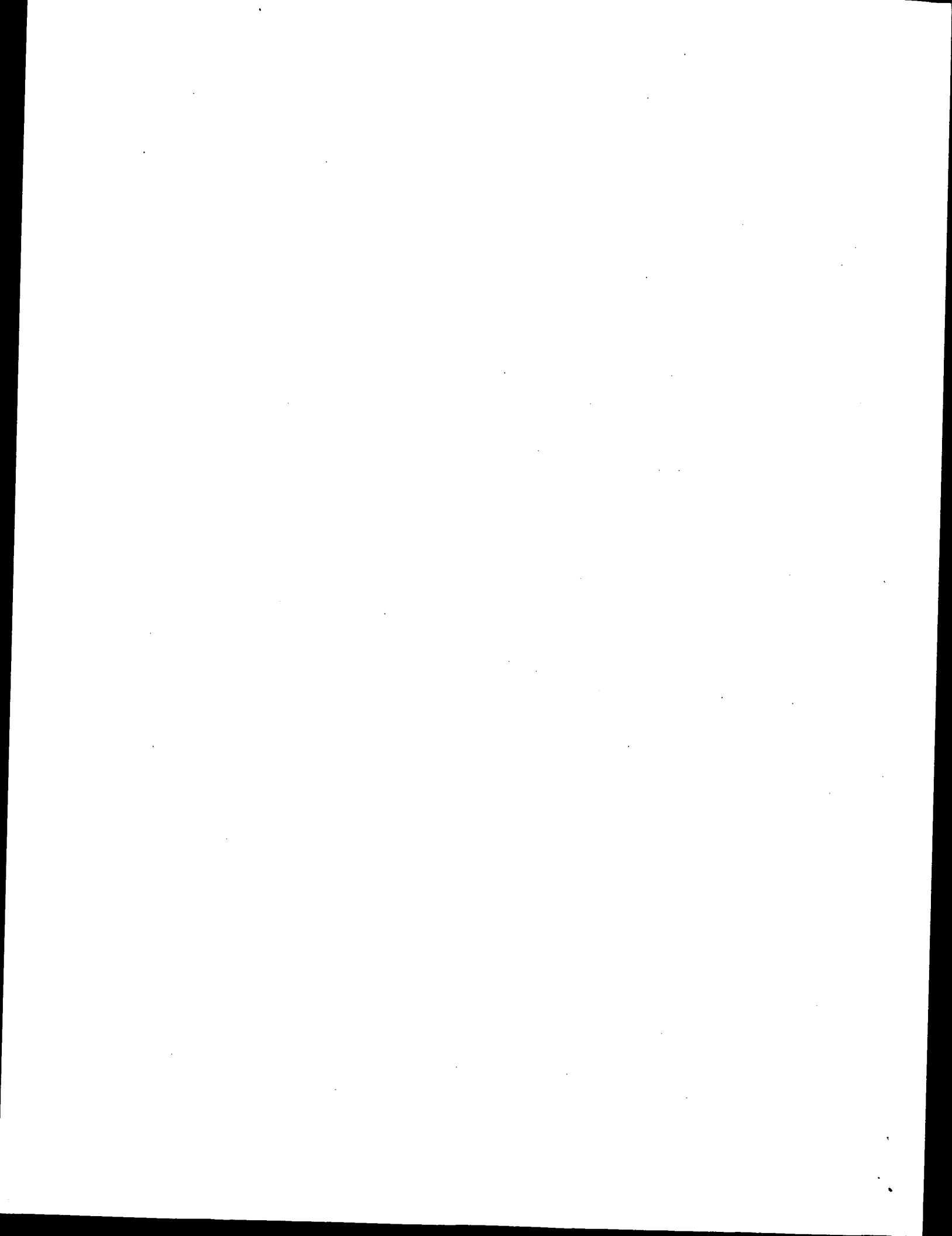
Search completed: May 11, 2003, 20:14:26  
Job time : 28.0354 secs

RESULT 44  
US-09-101-751A-74  
Sequence 74, Application US/09101751A  
Patent No. 6465253  
GENERAL INFORMATION:  
APPLICANT: WICKHAM, THOMAS J.  
APPLICANT: KOVESDI, IMRE  
APPLICANT: BROUGH, DOUGLAS E.  
TITLE OF INVENTION: VECTORS AND METHODS FOR GENE TRANSFER TO CELLS  
FILE REFERENCE: 85710  
CURRENT APPLICATION NUMBER: US/09/101,751A  
PRIOR FILING DATE: 1999-01-29  
CURRENT APPLICATION NUMBER: WO 96US19150  
PRIOR FILING DATE: 1996-11-27  
PRIOR APPLICATION NUMBER: US 08/700,846  
PRIOR FILING DATE: 1996-08-21  
PRIOR APPLICATION NUMBER: US 08/701,124  
PRIOR FILING DATE: 1996-08-21  
PRIOR APPLICATION NUMBER: US 08/563,368  
PRIOR FILING DATE: 1995-11-28  
NUMBER OF SEQ ID NOS: 94  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 74  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: ()..()  
OTHER INFORMATION: Description of Unknown Organism: Artificial  
OTHER INFORMATION: Sequence  
US-09-101-751A-74

Query Match 5.2%; Score 5; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKKR 31  
| | | | |  
Db 1 RKKKR 5

RESULT 45  
US-09-101-751A-94  
Sequence 94, Application US/09101751A  
Patent No. 6465253  
GENERAL INFORMATION:  
APPLICANT: WICKHAM, THOMAS J.  
APPLICANT: KOVESDI, IMRE  
APPLICANT: BROUGH, DOUGLAS E.  
TITLE OF INVENTION: VECTORS AND METHODS FOR GENE TRANSFER TO CELLS  
FILE REFERENCE: 85710  
CURRENT APPLICATION NUMBER: US/09/101,751A  
PRIOR FILING DATE: 1999-01-29  
CURRENT APPLICATION NUMBER: WO 96US19150  
PRIOR FILING DATE: 1996-11-27  
PRIOR APPLICATION NUMBER: US 08/700,846  
PRIOR FILING DATE: 1996-08-21  
PRIOR APPLICATION NUMBER: US 08/701,124  
PRIOR FILING DATE: 1996-08-21  
PRIOR APPLICATION NUMBER: US 08/563,368  
PRIOR FILING DATE: 1995-11-28  
NUMBER OF SEQ ID NOS: 94  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 94  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: ()..()  
OTHER INFORMATION: Sequence



Tue May 13 12:12:43 2003

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:12:57 ; Search time 31.7611 seconds  
(without alignments)  
281.051 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 97  
Sequence: 1 EVFVRDASLGDSITLSQT.....LTGCLPWPATRSHLGRKKS 97

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pcp.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pcp.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pcp.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	100.0	97	10	US-09-854-133-586
2	97	100.0	97	10	US-09-738-973-586
3	16	16.5	16	9	US-09-854-133-587
4	16	16.5	16	10	US-09-738-973-587
5	7	7.2	36	9	US-09-764-872-359
6	7	7.2	47	9	US-10-091-504-1018
7	7	7.2	47	10	US-09-764-869-1018
8	7	7.2	66	10	US-09-764-877-1089
9	7	7.2	94	9	US-09-866-050A-676
10	7	7.2	170	9	US-09-764-868-816
11	7	7.2	517	9	US-10-280-403-6
12	7	7.2	517	10	US-09-907-479-6
13	7	7.2	802	10	US-09-876-889-352
14	6	6.2	19	9	US-09-984-245-306
15	6	6.2	19	9	US-09-966-262-306
16	6	6.2	19	9	US-09-983-966-306
17	6	6.2	19	9	US-10-143-090-306
18	6	6.2	28	9	US-09-925-299-1023
19	6	6.2	28	10	US-09-925-299-1023

20	6	6.2	35	9	US-10-001-857-134	Sequence 134, Appl
21	6	6.2	35	12	US-10-001-870-145	Sequence 145, Appl
22	6	6.2	46	9	US-09-764-891-4700	Sequence 4700, Ap
23	6	6.2	51	10	US-09-864-761-37309	Sequence 37309, A
24	6	6.2	51	10	US-09-864-761-42719	Sequence 42719, A
25	6	6.2	59	10	US-09-864-761-43079	Sequence 43079, A
26	6	6.2	64	10	US-09-864-761-33889	Sequence 33889, A
27	6	6.2	69	10	US-09-864-761-34523	Sequence 34523, A
28	6	6.2	70	9	US-09-764-891-3256	Sequence 3256, Ap
29	6	6.2	73	10	US-09-250-611-19	Sequence 19, Appl
30	6	6.2	76	9	US-10-002-344A-229	Sequence 229, Appl
31	6	6.2	80	10	US-09-864-761-42011	Sequence 42011, A
32	6	6.2	89	10	US-10-072-349-99	Sequence 99, Appl
33	6	6.2	89	10	US-09-764-855-99	Sequence 99, Appl
34	6	6.2	97	10	US-10-091-504-1022	Sequence 1022, Ap
35	6	6.2	97	10	US-09-764-869-1022	Sequence 1022, Ap
36	6	6.2	101	9	US-09-925-299-774	Sequence 774, Appl
37	6	6.2	101	10	US-09-925-299-774	Sequence 774, Appl
38	6	6.2	123	9	US-10-043-487-333	Sequence 333, Appl
39	6	6.2	123	9	US-09-764-891-4621	Sequence 4621, Ap
40	6	6.2	128	10	US-09-864-761-36308	Sequence 36308, A
41	6	6.2	141	9	US-09-828-995B-29	Sequence 29, Appl
42	6	6.2	141	10	US-09-864-761-36181	Sequence 36181, A
43	6	6.2	151	10	US-09-789-561-88	Sequence 88, Appl
44	6	6.2	160	10	US-09-321-801-17	Sequence 17, Appl
45	6	6.2	163	9	US-10-211-962-77	Sequence 77, Appl
46	6	6.2	167	1	US-08-781-986A-5216	Sequence 5216, Ap
47	6	6.2	176	10	US-09-925-302-764	Sequence 764, Appl
48	6	6.2	184	10	US-09-250-611-13	Sequence 13, Appl
49	6	6.2	185	10	US-09-250-611-11	Sequence 11, Appl
50	6	6.2	230	10	US-09-925-300-1377	Sequence 1377, Ap
51	6	6.2	234	10	US-09-800-729-150	Sequence 150, Appl
52	6	6.2	246	9	US-10-157-223-5	Sequence 5, Appl
53	6	6.2	246	10	US-09-815-242-12986	Sequence 12986, A
54	6	6.2	246	10	US-09-815-242-13040	Sequence 13040, A
55	6	6.2	260	10	US-09-925-300-1470	Sequence 1470, Ap
56	6	6.2	261	9	US-09-738-626-6335	Sequence 6335, Ap
57	6	6.2	277	9	US-09-965-529-15	Sequence 15, Appl
58	6	6.2	281	9	US-10-101-464A-619	Sequence 619, Appl
59	6	6.2	290	10	US-09-910-174A-19	Sequence 19, Appl
60	6	6.2	318	10	US-09-947-971-2	Sequence 2, Appl
61	6	6.2	319	9	US-10-114-893-210	Sequence 210, Appl
62	6	6.2	319	12	US-10-025-335-1	Sequence 1, Appl
63	6	6.2	346	9	US-10-081-816-16	Sequence 16, Appl
64	6	6.2	350	10	US-09-910-174A-17	Sequence 17, Appl
65	6	6.2	352	9	US-09-828-995B-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1  
US-09-854-133-586  
; Sequence 586, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR LUNG CANCER  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 586  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-854-133-586

```
Query Match
Best Local Similarity 100.0%; Score 97; DB 9; Length 97;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVESRDHASLGDSEYLSQTELKRRKKRKKRKFQANGIDFIIFWIFWILLFHHWQ 60
DB 1 EVESRDHASLGDSEYLSQTELKRRKKRKKRKFQANGIDFIIFWIFWILLFHHWQ 60

QY 61 ESLLCPPSPKVTCTREMLTGGCLPWATRSHLGRKCS 97
DB 61 ESLLCPPSPKVTCTREMLTGGCLPWATRSHLGRKCS 97

RESULT 2
US-09-738-973-586
; Sequence 586, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 586
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-586

Query Match
Best Local Similarity 100.0%; Score 97; DB 10; Length 97;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVESRDHASLGDSEYLSQTELKRRKKRKKRKFQANGIDFIIFWIFWILLFHHWQ 60
DB 1 EVESRDHASLGDSEYLSQTELKRRKKRKKRKFQANGIDFIIFWIFWILLFHHWQ 60

QY 61 ESLLCPPSPKVTCTREMLTGGCLPWATRSHLGRKCS 97
DB 61 ESLLCPPSPKVTCTREMLTGGCLPWATRSHLGRKCS 97

RESULT 3
US-09-854-133-587
; Sequence 587, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 587
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-587

Query Match
Best Local Similarity 100.0%; Score 16; DB 9; Length 16;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 FOANGCIDFIIFWIFW 50
DB 1 FOANGCIDFIIFWIFW 16

RESULT 4
US-09-738-973-587
; Sequence 587, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 587
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-587

Query Match
Best Local Similarity 100.0%; Score 16; DB 10; Length 16;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 FOANGCIDFIIFWIFW 50
DB 1 FOANGCIDFIIFWIFW 16

RESULT 5
US-09-764-872-359
; Sequence 359, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 359
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-872-359
```



us-09-854-133-586.olig.rapb

Tue May 13 12:12:43 2003

FILE REFERENCE: PC005  
 CURRENT APPLICATION NUMBER: US/09/764,877  
 CURRENT FILING DATE: 2001-01-17  
 Prior application data removed - refer to PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 4031  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 1089  
 LENGTH: 66  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-764-877-1089

Query Match 7.2%; Score 7; DB 10; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 9.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30  
 Db 35 KKERKK 41

RESULT 9

US-09-866-050A-676  
 Sequence 676, Application US/09866050A  
 Publication No. US20030040471A1  
 GENERAL INFORMATION:  
 APPLICANT: Watson, James D.  
 APPLICANT: Strachan, Lorna  
 APPLICANT: Sleeman, Matthew  
 APPLICANT: Onrust, Rene  
 APPLICANT: Murison, James G.  
 APPLICANT: Kumble, Krishanand D.  
 TITLE OF INVENTION: Compositions Isolated From Skin Cells  
 TITLE OF INVENTION: and Methods for Their Use  
 FILE REFERENCE: 11000.1011c4U  
 CURRENT APPLICATION NUMBER: US/09/866,050A  
 CURRENT FILING DATE: 2001-05-24  
 NUMBER OF SEQ ID NOS: 725  
 SOFTWARE: FastSEQ for Windows Version 4.0  
 SEQ ID NO 676  
 LENGTH: 94  
 TYPE: PRT  
 ORGANISM: Mouse  
 US-09-866-050A-676

Query Match 7.2%; Score 7; DB 9; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29  
 Db 35 RKKERKK 41

RESULT 10

US-09-764-868-816  
 Sequence 816, Application US/09764868  
 Patent No. US20020168711A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PT232  
 CURRENT APPLICATION NUMBER: US/09/764,868  
 CURRENT FILING DATE: 2001-01-17  
 Prior application data removed - refer to PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 1510  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 816  
 LENGTH: 170  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 NAME/KEY: SITE

Query Match 7.2%; Score 7; DB 9; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 5.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30  
 Db 7 KKERKK 13

RESULT 6

US-10-091-504-1018  
 Sequence 1018, Application US/10091504  
 Publication No. US2003005908A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PC007C1  
 CURRENT APPLICATION NUMBER: US/10/091,504  
 CURRENT FILING DATE: 2002-03-07  
 NUMBER OF SEQ ID NOS: 2442  
 Prior application data removed - See File Wrapper or Palm  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 1018  
 LENGTH: 47  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-091-504-1018

Query Match 7.2%; Score 7; DB 9; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KKERKK 31  
 Db 14 KKERKK 20

RESULT 7

US-09-764-869-1018  
 Sequence 1018, Application US/09764869  
 Patent No. US20020061521A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PC007  
 CURRENT APPLICATION NUMBER: US/09/764,869  
 CURRENT FILING DATE: 2001-01-17  
 Prior application data removed - refer to PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 2442  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 1018  
 LENGTH: 47  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-764-869-1018

Query Match 7.2%; Score 7; DB 10; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KKERKK 31  
 Db 14 KKERKK 20

RESULT 8

US-09-764-877-1089  
 Sequence 1089, Application US/09764877  
 Patent No. US20020147140A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; LOCATION: (118)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; US-09-764-868-816

Query Match 7.2%; Score 7; DB 9; Length 170;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKEKK 29  
| | | | |  
DB 17 RKKEKK 23

## RESULT 11

US-10-280-403-6  
; Sequence 6, Application US/10280403  
; Publication No. US20030082620A1  
; GENERAL INFORMATION:

; APPLICANT: Astle, Jon H  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Dwivedi, Poorima  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Molino, Gary A.  
; APPLICANT: Myerow, Susan H.  
; APPLICANT: Thiagalingam, Arunthathi  
; APPLICANT: Catino, Theodore  
; TITLE OF INVENTION: No. US20030082620A1el Human Genes and Gene Expression Products:  
; FILE REFERENCE: 1657/1015B  
; CURRENT APPLICATION NUMBER: US/10/280,403  
; CURRENT FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: US 09/385,982  
; PRIOR FILING DATE: 1999-08-30  
; PRIOR APPLICATION NUMBER: US 09/328,111  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: US 60/098,639  
; PRIOR FILING DATE: 1988-08-31  
; PRIOR APPLICATION NUMBER: US 60/117,393  
; PRIOR FILING DATE: 1998-01-27  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-280-403-6

Query Match 7.2%; Score 7; DB 9; Length 517;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31  
| | | | |  
DB 300 KERKKR 306

## RESULT 12

US-09-907-479-6  
; Sequence 6, Application US/09907479  
; Patent No. US20020034758A1  
; GENERAL INFORMATION:

; APPLICANT: Astle, Jon  
; APPLICANT: Burgess, Christopher  
; APPLICANT: Dwivedi, Poorima  
; APPLICANT: Lewis, Marcia  
; APPLICANT: Molino, Gary  
; APPLICANT: Myerow, Susan  
; APPLICANT: Thiagalingam, Arunthathi  
; APPLICANT: Catino, Theodore  
; TITLE OF INVENTION: No. US20020034758A1el Human Genes and Gene Expression Products:  
; FILE REFERENCE: 1657/1015B  
; CURRENT APPLICATION NUMBER: US/09/907,479  
; CURRENT FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: US 09/385,982

; PRIOR FILING DATE: 1999-08-30  
; PRIOR APPLICATION NUMBER: US 09/328,111  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: US 60/098,639  
; PRIOR FILING DATE: 1988-08-31  
; PRIOR APPLICATION NUMBER: US 60/117,393  
; PRIOR FILING DATE: 1998-01-27  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-907-479-6

Query Match 7.2%; Score 7; DB 10; Length 517;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31  
| | | | |  
DB 300 KERKKR 306

## RESULT 13

US-09-876-889-352  
; Sequence 352, Application US/09876889  
; Patent No. US20020076715A1  
; GENERAL INFORMATION:

; APPLICANT: Benson, Darin R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: King, Gordon E.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN  
; FILE REFERENCE: 210121.466C3  
; CURRENT APPLICATION NUMBER: US/09/876,889  
; CURRENT FILING DATE: 2001-06-06  
; NUMBER OF SEQ ID NOS: 353  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 352  
; LENGTH: 802  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-09-876-889-352

Query Match 7.2%; Score 7; DB 10; Length 802;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SQTELK 24  
| | | | |  
DB 282 SQTELK 288

## RESULT 14

US-09-984-245-306  
; Sequence 306, Application US/09984245  
; Patent No. US20020165374A1  
; GENERAL INFORMATION:

; APPLICANT: Young et al.  
; TITLE OF INVENTION: 87 Human Secreted Proteins  
; FILE REFERENCE: P2004P1  
; CURRENT APPLICATION NUMBER: US/09/984,245  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 09/154,707  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: PC7/US98/05311  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: US 60/041,277  
; PRIOR FILING DATE: 1997-03-21  
; PRIOR APPLICATION NUMBER: US 60/042,344  
; PRIOR FILING DATE: 1997-03-21

us-09-854-133-586.olig.rapb

Tue May 13 12:12:43 2003

```
; PRIOR APPLICATION NUMBER: US 60/041,276
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,281
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/048,094
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,350
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,188
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,135
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,094
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/050,937
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,187
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,099
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,352
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,131
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,096
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,355
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,160
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,351
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,154
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/054,804
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: US 60/056,370
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: US 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 343
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 306
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-245-306

Query Match      6.2%; Score 6; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 KKERK 29
Db 10 KKERK 15

RESULT 15
US-09-966-262-306
; Sequence 306, Application US/09966262
; Publication No. US20030050461A1
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 87 Human Secreted Proteins
; FILE REFERENCE: P2004P1
; CURRENT APPLICATION NUMBER: US/09/966,262
; CURRENT FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: US 09/154,707
; PRIOR FILING DATE: 1998-09-17
```

```
; PRIOR APPLICATION NUMBER: PCT/US98/05311
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: US 60/041,277
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/042,344
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,276
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,281
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/048,094
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,350
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,188
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,135
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/050,937
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,187
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,099
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,352
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,186
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,069
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,095
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,131
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,096
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,355
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,160
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,351
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,154
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/054,804
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: US 60/056,370
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: US 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 343
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 306
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-262-306

Query Match      6.2%; Score 6; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 KKERK 29
Db 10 KKERK 15

RESULT 16
US-09-983-966-306
; Sequence 306, Application US/09983966
; Publication No. US20030060619A1
; GENERAL INFORMATION:
; APPLICANT: Young et al.
```

RESULT 17  
US-10-143-090-306  
; Sequence 306, Application US/10143090  
; Publication No. US20030069406A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 87 Human Secreted Proteins  
; FILE REFERENCE: P2004P1  
; CURRENT APPLICATION NUMBER: US/10/143,090  
; CURRENT FILING DATE: 2002-05-13  
; PRIOR APPLICATION NUMBER: 09/154,707  
; PRIOR FILING DATE: 1998-09-17  
; NUMBER OF SEQ ID NOS: 343  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 306  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-143-090-306

Query Match 6.2%; Score 6; DB 9; Length 19;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 29  
Db 10 KKERKK 15

RESULT 18  
US-09-925-299-1023  
; Sequence 1023, Application US/09925299  
; Publication No. US20030040617A9  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1023  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-299-1023

Query Match 6.2%; Score 6; DB 9; Length 28;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
Db 14 SETLSQ 19

RESULT 19  
US-09-925-299-1023  
; Sequence 1023, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270

RESULT 17  
US-10-143-090-306  
; Sequence 306, Application US/10143090  
; Publication No. US20030069406A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 87 Human Secreted Proteins  
; FILE REFERENCE: P2004P1  
; CURRENT APPLICATION NUMBER: US/10/143,090  
; CURRENT FILING DATE: 2002-05-13  
; PRIOR APPLICATION NUMBER: 09/154,707  
; PRIOR FILING DATE: 1998-09-17  
; NUMBER OF SEQ ID NOS: 343  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 306  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-143-090-306

Query Match 6.2%; Score 6; DB 9; Length 19;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 29  
Db 10 KKERKK 15

RESULT 18  
US-09-925-299-1023  
; Sequence 1023, Application US/09925299  
; Publication No. US20030040617A9  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1023  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-299-1023

Query Match 6.2%; Score 6; DB 9; Length 28;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19  
Db 14 SETLSQ 19

RESULT 19  
US-09-925-299-1023  
; Sequence 1023, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270

us-09-854-133-586.olig.rapb

Tue May 13 12:12:43 2003

```

; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1023
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1023

Query Match
Best Local Similarity 100.0%; Pred. No. 43; Length 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19
DB 14 SETLSQ 19

RESULT 20
US-10-001-857-134
; Sequence 134, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 134
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-857-134

Query Match
Best Local Similarity 100.0%; Pred. No. 52; Length 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19
DB 20 SETLSQ 25

RESULT 21
US-10-001-870-145
; Sequence 145, Application US/10001870
; Patent No. US20020150924A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and Proteins
; FILE REFERENCE: DEX-0283
; CURRENT APPLICATION NUMBER: US/10/001,870
; CURRENT FILING DATE: 2001-11-20
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,189
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 145
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapien

US-10-001-870-145

Query Match
Best Local Similarity 100.0%; Pred. No. 52; Length 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19
DB 20 SETLSQ 25

RESULT 22
US-09-764-891-4700
; Sequence 4700, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4700
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (28)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4700

Query Match
Best Local Similarity 100.0%; Pred. No. 66; Length 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SETLSQ 19
DB 35 SETLSQ 40

RESULT 23
US-09-864-761-37309
; Sequence 37309, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37309
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010885.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
US-09-864-761-37309

```

```

Query Match          6.2%; Score 6; DB 10; Length 51;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 29 KKRRK 34
Db 4 KKRRK 9

```

```

RESULT 24
US-09-864-761-42719
; Sequence 42719, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42719
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010885.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
US-09-864-761-42719

```

```

Query Match          6.2%; Score 6; DB 10; Length 51;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 29 KKRRK 34
Db 4 KKRRK 9

```

```

RESULT 25
US-09-864-761-43079
; Sequence 43079, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30

```

```

Query Match          6.2%; Score 6; DB 10; Length 59;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 0; Indels
QY      27 RKKKRE 32
          ||| |||
Db       6 RKKKRE 11

```

US-09-864-761-33889  
; sequence 33889, Application US/09864761  
: Patent No US20020048763A1

```

1  Patent No. 00/000000
2  GENERAL INFORMATION:
3  APPLICANT: Penn, Sharron G.
4  APPLICANT: Rank, David R.
5  APPLICANT: Chan, David K.
6  APPLICANT: Hzen, Wensheng
7  TITLE OF INVENTION: HUMAN GENOME-DERIVED
8  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS
9  FILE REFERENCE: Aesomica-X-1
10 CURRENT APPLICATION NUMBER: US/09/864,
11 CURRENT FILING DATE: 2001-05-23
12 PRIOR APPLICATION NUMBER: US 60/180,311
13 PRIOR FILING DATE: 2000-02-04
14 PRIOR APPLICATION NUMBER: US 60/207,451
15 PRIOR FILING DATE: 2000-05-26
16 PRIOR APPLICATION NUMBER: US 09/632,361
17 PRIOR FILING DATE: 2000-08-03
18 PRIOR APPLICATION NUMBER: GB 24263.6
19 PRIOR FILING DATE: 2000-10-04
20 PRIOR APPLICATION NUMBER: US 60/236,351
21 PRIOR FILING DATE: 2000-09-27

```

APPLICANT: Hanzel, David K.  
 APPLICANT: Chen, Wensheng  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
 FILE REFERENCE: Aecomica-x-1  
 CURRENT APPLICATION NUMBER: US/09/864,761  
 CURRENT FILING DATE: 2001-05-23  
 PRIOR APPLICATION NUMBER: US 60/180,312  
 PRIOR FILING DATE: 2000-02-04  
 PRIOR APPLICATION NUMBER: US 60/207,456  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: US 09/632,366  
 PRIOR FILING DATE: 2000-08-03  
 PRIOR APPLICATION NUMBER: GB 24263.6  
 PRIOR FILING DATE: 2000-10-04  
 PRIOR APPLICATION NUMBER: US 60/236,359  
 PRIOR FILING DATE: 2000-09-27

QY	59 IQESLL	64
pb	2 IQESLL	7

```

RESULT 27
US-09-864-761-34523
,, Sequence 34523, Application US/09864761
,, Patent No. US20020048763A1
,, GENERAL INFORMATION:
,, APPLICANT: Penn, Sharon G.
,, APPLICANT: Rank, David R.
,, APPLICANT: Hanzel, David K.
,, APPLICANT: Chen, Wensheng
,, TITLE OF INVENTION: HUMAN GENOME-DERIVED
,, TITLE OF INVENTION: GENE EXPRESSION AN
,, FILE REFERENCE: Aeomica-x-1
,, CURRENT APPLICATION NUMBER: US/09/864,7
,, CURRENT FILING DATE: 2001-05-23
,, PRIOR APPLICATION NUMBER: US 60/180,312
,, PRIOR FILING DATE: 2000-02-04

```

```
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34523
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006028.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 14
; OTHER INFORMATION: EST_HUMAN HIT: BE071726.1, EVALUATE 5.40e-01
US-09-864-761-34523
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 10; Length 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 25 KERKK 30
    |||||
DB 36 KERKK 41
```

```
RESULT 28
US-09-764-891-3256
; Publication 3256, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
```

```
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3256
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (15)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3256
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 9; Length 70;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 28 KKKRR 33
    |||||
DB 1 KKKRR 6
```

```
RESULT 29
US-09-250-611-19
; Sequence 19, Application US/09250611
; Patent No. US20020143161A1
; GENERAL INFORMATION:
; APPLICANT: Byrnie, Paul
; APPLICANT: Basset, Paul
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210001
; CURRENT APPLICATION NUMBER: US/09/250,611
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-250-611-19
```

```
Query Match
Best Local Similarity 6.2%; Score 6; DB 10; Length 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 14 SETLSQ 19
    |||||
DB 16 SETLSQ 21
```

```
RESULT 30
US-10-002-344A-229
; Sequence 229, Application US/10002344A
; Patent No. US20020172959A1
; GENERAL INFORMATION:
; APPLICANT: Recipon, Herve
; APPLICANT: Sun, Yongming
; APPLICANT: Chen, Sei-yu
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and
; FILE REFERENCE: DEX-0241
; CURRENT APPLICATION NUMBER: US/10/002,344A
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/242,998
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 277
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 229
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapien
```



us-09-854-133-586.olig.rapb

Tue May 13 12:12:43 2003

```

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 9.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 8.5
; OTHER INFORMATION: EST_HUMAN HIT: AV718492.1, EVALUATE 3.90e+00
US-09-864-761-42011

```

```

Query Match      6.2%; Score 6; DB 10; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Indels
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

```

```

QY 24 KKERKK 29
DB 44 KKERKK 49

```

```

RESULT 32
US-10-072-349-99
; Sequence 99, Application US/10072349
; Publication No. US20030054420A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1110C1
; CURRENT APPLICATION NUMBER: US/10/072,349
; CURRENT FILING DATE: 2002-02-11
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 99
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-349-99

```

```

Query Match      6.2%; Score 6; DB 9; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

```

```

QY 14 SETLSQ 19
DB 45 SETLSQ 50

```

```

RESULT 33
US-09-764-855-99
; Sequence 99, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 99
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-855-99

```

```

Query Match      6.2%; Score 6; DB 10; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

```

```

QY 14 SETLSQ 19
DB 45 SETLSQ 50

```

RESULT 34

US-10-002-344A-229

```

Query Match      6.2%; Score 6; DB 9; Length 76;
Best Local Similarity 100.0%; Pred. No. 1e+02; Indels
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

```

```

QY 24 KKERKK 29
DB 62 KKERKK 67

```

```

RESULT 31
US-09-864-761-42011
; Sequence 42011, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-08-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42011
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007340.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.8

```

US-10-091-504-1022  
; Sequence 1022, Application US/10091504  
; Publication No. US20030059908A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007C1  
; CURRENT APPLICATION NUMBER: US/10/091,504  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2442  
; Prior Application removed - See File Wrapper or Palm  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1022  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-091-504-1022

Query Match  
Best Local Similarity 6.2%; Score 6; DB 9; Length 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKK 25  
Db 74 TELRKK 79

## RESULT 35

US-09-764-869-1022  
; Sequence 1022, Application US/09764869  
; Patent No. US20030061521A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007  
; CURRENT APPLICATION NUMBER: US/09/764,869  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2442  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1022  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-869-1022

Query Match  
Best Local Similarity 6.2%; Score 6; DB 10; Length 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKK 25  
Db 74 TELRKK 79

## RESULT 36

US-09-925-299-774  
; Sequence 774, Application US/09925299  
; Publication No. US20030040817A9  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 774  
; LENGTH: 101

; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (69)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (98)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-774

Query Match  
Best Local Similarity 6.2%; Score 6; DB 9; Length 101;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPKE 71  
Db 52 PPSPKE 57

## RESULT 37

US-09-925-299-774  
; Sequence 774, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 774  
; LENGTH: 101  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (69)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (98)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-774

Query Match  
Best Local Similarity 6.2%; Score 6; DB 10; Length 101;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPKE 71  
Db 52 PPSPKE 57

## RESULT 38

US-10-043-487-333  
; Sequence 333, Application US/10043487  
; Publication No. US20030055220A1  
; GENERAL INFORMATION:  
; APPLICANT: HYBRIGENICS  
; APPLICANT: Pierre, LEGRAIN  
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptides  
; FILE REFERENCE: B4778A  
; CURRENT APPLICATION NUMBER: US/10/043,487  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/261,130  
; PRIOR FILING DATE: 2001-01-12  
; NUMBER OF SEQ ID NOS: 561



; NUMBER OF SEQ ID NOS: 104  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 29  
 ; LENGTH: 141  
 ; TYPE: PRT  
 ; ORGANISM: Canis familiaris  
 US-09-828-995B-29

Query Match  
 Best Local Similarity 6.2%; Score 6; DB 9; Length 141;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 66 PPSPE 71  
 Db 43 PPSPE 48

RESULT 42  
 US-09-864-761-36181  
 ; Sequence 36181, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Aeonica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 36181  
 ; LENGTH: 141  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens

; FEATURE:  
 ; OTHER INFORMATION: MAP TO AL022333.1  
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4  
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4  
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.3  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3  
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.2  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9  
 US-09-864-761-36181

Query Match  
 Best Local Similarity 6.2%; Score 6; DB 10; Length 141;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERK 28  
 Db 44 RKKERK 49

RESULT 43  
 US-09-789-561-88  
 ; Sequence 88, Application US/09789561  
 ; Patent No. US20020064818A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ni et al.  
 ; TITLE OF INVENTION: 52 Human secreted proteins  
 ; FILE REFERENCE: P2043P1  
 ; CURRENT APPLICATION NUMBER: US/09/789,561  
 ; CURRENT FILING DATE: 2001-02-22  
 ; PRIOR APPLICATION NUMBER: PCT/US00/24008  
 ; PRIOR FILING DATE: 2000-08-31  
 ; PRIOR APPLICATION NUMBER: 60/152,317  
 ; PRIOR FILING DATE: 1999-09-03  
 ; PRIOR APPLICATION NUMBER: 60/152,315  
 ; PRIOR FILING DATE: 1999-09-03  
 ; NUMBER OF SEQ ID NOS: 194  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 88  
 ; LENGTH: 151  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-789-561-88

Query Match  
 Best Local Similarity 6.2%; Score 6; DB 10; Length 151;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 63 LLCPPS 68  
 Db 143 LLCPPS 148

RESULT 44  
 US-09-321-801-17  
 ; Sequence 17, Application US/09321801  
 ; Patent No. US20020115176A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lanfear, Jeremy  
 ; APPLICANT: Robas, Nicola M.  
 ; TITLE OF INVENTION: PHOSPHODIESTERASE ENZYMES  
 ; FILE REFERENCE: EC9477A  
 ; CURRENT APPLICATION NUMBER: US/09/321,801  
 ; CURRENT FILING DATE: 1999-05-27  
 ; EARLIER APPLICATION NUMBER: 9826777.6  
 ; EARLIER FILING DATE: 1998-12-04  
 ; EARLIER APPLICATION NUMBER: 9823882.7  
 ; EARLIER FILING DATE: 1998-10-30  
 ; EARLIER APPLICATION NUMBER: 9811500.9  
 ; EARLIER FILING DATE: 1998-05-28

us-09-854-133-586.olig.rapb

Tue May 13 12:12:43 2003

EARLIER APPLICATION NUMBER: 9908247.1  
 EARLIER FILING DATE: 1998-04-09  
 EARLIER APPLICATION NUMBER: 9910801.1  
 EARLIER FILING DATE: 1999-05-10  
 NUMBER OF SEQ ID NOS: 19  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 17  
 LENGTH: 160  
 TYPE: PRT  
 ORGANISM: Rat  
 US-09-321-801-17

Query Match 6.2%; Score 6; DB 10; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 PSPKEV 72  
 Db 68 PSPKEV 73

# RESULT 45

US-10-211-962-77  
 Sequence 77, Application US/10211962  
 Publication No. US20030082640A1  
 GENERAL INFORMATION:  
 APPLICANT: Herz, Joachim  
 APPLICANT: Gotthardt, Michael  
 TITLE OF INVENTION: LDL Receptor Signaling Pathways  
 FILE REFERENCE: UTSW0708  
 CURRENT APPLICATION NUMBER: US/10/211,962  
 CURRENT FILING DATE: 2002-08-01  
 PRIOR APPLICATION NUMBER: US/09/562,737  
 PRIOR FILING DATE: 2000-05-01  
 NUMBER OF SEQ ID NOS: 132  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 77  
 LENGTH: 163  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 OTHER INFORMATION: Sequence  
 US-10-211-962-77

Query Match 6.2%; Score 6; DB 9; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGDS 14  
 Db 13 ASLGDS 18

Search completed: May 11, 2003, 20:21:13  
 Job time : 34.7611 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:07:52 ; Search time 30.042 Seconds  
(without alignments)  
310.377 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 97  
Sequence: 1 EVEVSRDHASLGDSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : PIR.73:.\*  
1: Pirl:.\*  
2: Pirl2:.\*  
3: Pirl3:.\*  
4: Pirl4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	8	8.2	67	A70375	Ribosomal protein
2	7	7.2	129	B71554	Ribosomal protein
3	7	7.2	133	H81705	Ribosomal protein
4	7	7.2	134	B72101	Ribosomal protein
5	7	7.2	134	F86521	Ribosomal protein
6	7	7.2	197	T17106	hypothetical prote
7	7	7.2	310	E64923	probable transcrip
8	7	7.2	310	H90924	probable transcrip
9	7	7.2	310	D85773	probable transcrip
10	7	7.2	310	A10695	hypothetical prote
11	7	7.2	318	C84651	hypothetical prote
12	7	7.2	340	C75004	hypothetical prote
13	7	7.2	356	S55181	conserved hypothe
14	7	7.2	383	F82871	probable extensin
15	7	7.2	394	C84905	probable allantoin
16	7	7.2	409	AF3271	hypothetical prote
17	7	7.2	463	AH2025	cytochrome-c oxida
18	7	7.2	483	S41689	hypothetical prote
19	7	7.2	483	F71619	hypothetical prote
20	7	7.2	557	T04465	hypothetical prote
21	7	7.2	560	F70719	cysteine proteinas
22	7	7.2	583	S46265	hypothetical prote
23	7	7.2	628	T01467	DNA helicase homol
24	7	7.2	648	C69423	hypothetical prote
25	7	7.2	651	C86333	modulation protein
26	7	7.2	703	C75638	hypothetical prote
27	7	7.2	777	T04768	protein 4.1, P4.1
28	7	7.2	858	A46613	hypothetical prote
29	7	7.2	1257	T28937	

30	7	7.2	1286	2	T16507	hypothetical prote
31	7	7.2	1435	2	S69632	regulatory protein
32	7	7.2	1551	2	F86342	F9H16.4 protein -
33	7	7.2	3844	2	T18402	asparagine/asparta
34	6	6.2	63	2	D69398	hypothetical prote
35	6	6.2	74	2	AG3625	hypothetical prote
36	6	6.2	80	2	G84773	hypothetical prote
37	6	6.2	89	2	B84145	smt3 protein - ric
38	6	6.2	100	2	T04102	hypothetical prote
39	6	6.2	103	2	T17566	cold stress protei
40	6	6.2	110	2	T07618	transcription repr
41	6	6.2	116	2	B40617	mazg-related prote
42	6	6.2	116	2	H72386	transcription regu
43	6	6.2	121	2	D72336	hypothetical prote
44	6	6.2	121	2	F84868	hypothetical prote
45	6	6.2	124	2	T37059	ribosome-binding f
46	6	6.2	129	2	F83052	probable signal re
47	6	6.2	132	2	T04105	conserved hypothe
48	6	6.2	132	2	B69454	ribosome-binding f
49	6	6.2	136	2	B82298	Ribosomal protein
50	6	6.2	137	2	S73212	bZIP transcription
51	6	6.2	142	2	T46042	30S ribosomal prot
52	6	6.2	143	2	T46122	hypothetical prote
53	6	6.2	144	2	T18867	N-terminal acetyl
54	6	6.2	156	2	G69233	hypothetical prote
55	6	6.2	162	2	H70314	hypothetical prote
56	6	6.2	162	2	D37844	baIE 19.5K protein
57	6	6.2	167	2	T10639	hypothetical prote
58	6	6.2	169	2	G75319	hypothetical prote
59	6	6.2	173	2	T45993	hypothetical 19.8K
60	6	6.2	184	2	I38910	pulmonary surfacta
61	6	6.2	189	1	LNRC1	hypothetical prote
62	6	6.2	201	2	G70156	hypothetical prote
63	6	6.2	205	2	E64621	conserved hypothe
64	6	6.2	205	2	T00865	hypothetical prote
65	6	6.2	205	2	C70950	hypothetical prote

## ALIGNMENTS

RESULT 1  
A70375 Ribosomal protein S21 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 13-Aug-1999  
C:Accession: A70375  
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: A70375  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-67 <AQF>  
A:Cross-references: GB:AE000711; NID:g2983401; PIDN:AC06990.1; PID:g2983406; GB:AE000711  
A:Experimental source: strain VF5  
C:Genetics:  
C:Superfamily: Escherichia coli ribosomal protein S21

Query Match 8.2%; Score 8; DB 2; Length 67;  
Best Local Similarity 100.0%; Pred. No. 0.33;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 26 ERKKKER 33  
Db 43 ERKKKER 50

RESULT 2  
B71554

ribosomal protein S9 [similarity] - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
 C:Species: Chlamydia trachomatis  
 C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 03-May-2002  
 C:Accession: B71554  
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998  
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis  
 A:Reference number: A71570; MUID:99000809; PMID:9784136  
 A:Accession: B71554  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-129 <ARN>  
 A:Cross-references: GB:AE001286; GB:AE001273; NID:g3328516; PIDN:AAC67717.1; PID:g332852  
 A:Experimental source: serotype D, strain UW-3/Cx  
 C:Genetics:  
 A:Gene: rs9  
 C:Superfamily: Escherichia coli ribosomal protein S9

Query Match 7.2%; Score 7; DB 2; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 6.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29  
 |||||  
 Db 108 RKKERKK 114

RESULT 3  
 H81705  
 ribosomal protein S9 TC0402 [imported] - Chlamydia muridarum (strain Nigg)  
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
 C:Accession: H81705  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; MUID:20150255; PMID:10684935  
 A:Accession: H81705  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-133 <TET>  
 A:Cross-references: GB:AE002307; GB:AE002160; NID:g7190442; PIDN:AAF39259.1; PID:g719044  
 A:Experimental source: strain Nigg (MoPn)  
 C:Genetics:  
 A:Gene: TC0402  
 C:Superfamily: Escherichia coli ribosomal protein S9

Query Match 7.2%; Score 7; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29  
 |||||  
 Db 112 RKKERKK 118

RESULT 4  
 B72101  
 ribosomal protein S9 CP0516 [similarity] - Chlamydia pneumoniae (strains CWL029 and C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 03-May-2002  
 C:Accession: B72101; F81567  
 R:Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999  
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; MUID:99206606; PMID:10192388  
 A:Accession: B72101  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <ARN>  
 A:Cross-references: GB:AE001610; GB:AE001363; NID:g4376515; PIDN:RAD18399.1; PID:g437652  
 A:Experimental source: strain CWL029  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Sal Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae A  
 A:Reference number: A81500; MUID:20150255; PMID:10684935  
 A:Accession: F81567  
 A:Molecule type: DNA  
 A:Residues: 1-134 <REA>  
 A:Cross-references: GB:AE002212; GB:AE002161; NID:g7189430; PIDN:AAF38343.1; PID:g  
 A:Experimental source: strain AR39, HL cells  
 C:Genetics:  
 A:Gene: rs9; CP0516  
 C:Superfamily: Escherichia coli ribosomal protein S9

Query Match 7.2%; Score 7; DB 2; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29  
 |||||  
 Db 113 RKKERKK 119

RESULT 5  
 F86521  
 ribosomal protein S9 [similarity] - Chlamydia pneumoniae (strain J138)  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 03-May-2002  
 C:Accession: F86521  
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T. Nucleic Acids Res. 28, 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.  
 A:Reference number: A86491; MUID:20330349; PMID:10871362  
 A:Accession: F86521  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <STO>  
 A:Cross-references: GB:BA000008; NID:g8978619; PIDN:BAA98456.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: rs9  
 C:Superfamily: Escherichia coli ribosomal protein S9

Query Match 7.2%; Score 7; DB 2; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKERKK 29  
 |||||  
 Db 113 RKKERKK 119

RESULT 6  
 T17106  
 hypothetical protein PAFD103 - apple tree (fragment)  
 C:Species: Malus domestica (apple tree)  
 C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
 C:Accession: T17106  
 R:Dong, Y.H.; Janssen, B.J.; Bielecki, L.L.; Atkinson, R.G.; Morris, B.A.; Gardner, J. Am. Soc. Hort. Sci. 122, 752-757, 1997  
 A:Title: Isolating and characterizing genes differentially expressed early in apple  
 A:Reference number: Z18681  
 A:Accession: T17106  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-197 <DON>  
 A:Cross-references: EMBL:U80270; NID:gl732362; PID:gl732363  
 A:Experimental source: strain Granny Smith  
 C:Genetics:  
 A:Note: PAFD103

Query Match 7.2%; Score 7; DB 2; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 9.6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: D85773

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimaanta, E.; Potamousis, K.; Apoc

Nature 409, 529-533, 2001

A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: D85773

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-310 <STO>

A:Cross-references: GB:AE005174; NID:gl2515657; PIDN:AAG56648.1; GSPDB:GN00145; UWG

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ydhB

C:Superfamily: probable transcription regulator ybbs

Query Match 7.2%; Score 7; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17

Db 288 LGDSETL 294

RESULT 10

AI0895

probable transcription regulator SRY1693 [imported] - Salmonella enterica subsp. en

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AI0695

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Chur

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fa

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens

A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica

A:Reference number: AB0502; PMID:11677608

A:Accession: AI0695

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-310 <PAK>

A:Cross-references: GB:AL513382; PIDN:CAD01938.1; PID:gl6502780; GSPDB:GN00176

C:Genetics:

A:Gene: STY1693

C:Superfamily: probable transcription regulator ybbs

Query Match 7.2%; Score 7; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17

Db 288 LGDSETL 294

RESULT 11

C84651

hypothetical protein At2g25670 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: C84651

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii,

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tall

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ven

Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10517197

A:Accession: C84651

A>Status: preliminary

A:Molecule type: DNA

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Mar-2002

C:Accession: E64923

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

Science 277, 1453-1462, 1997

A>Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: E64923

A:Molecule type: DNA

A:Residues: 1-310 <BLAT>

A:Cross-references: GB:AE000261; GB:U00096; NID:gl787945; PIDN:AAC74731.1; PID:gl787949;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: ydhB

C:Superfamily: probable transcription regulator ybbs

C:Keywords: DNA binding; transcription regulation

F:18-48/Region: regulatory protein lysR motif

F:19-38/Region: helix-turn-helix motif

Query Match 7.2%; Score 7; DB 1; Length 310;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17

Db 288 LGDSETL 294

RESULT 8

H90924

probable transcription regulator LYSR-type [imported] - Escherichia coli (strain O157:H7

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C:Accession: H90924

R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: H90924

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-310 <HAY>

A:Cross-references: PIDN:BA000007; PIDN:BA035791.1; PID:gl3361835; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECS2368

C:Superfamily: probable transcription regulator ybbs

Query Match 7.2%; Score 7; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17

Db 288 LGDSETL 294

RESULT 9

D85773

probable transcription regulator LYSR-type ydhB [imported] - Escherichia coli (strain O1

C:Species: Escherichia coli

A:Residues: 1-318 <STO>  
A:Cross-References: GB:AE002093; NID:g4874305; PIDN:AAD31367.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g25670  
A:Map position: 2

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 318;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 KKKKKK 30  
|||||||  
Db 160 KKKKKK 166

RESULT 12  
C75004  
hypothetical protein PAB1036 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: C75004  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure  
A:Reference number: A75001  
A:Accession: C75004  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-340 <KAF>  
A:Cross-References: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50473.1; PID:g545898  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB1036  
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB1036

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 340;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 KKKKKR 31  
|||||||  
Db 287 KKKKKR 293

RESULT 13  
S55181  
hypothetical protein YJL131c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein J0682  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002  
C:Accession: S55181; S05768; S56913; S71666; PQ0003  
R:Katsoulou, C.; Tzermia, M.; Alexandraki, D.  
submitted to the EMBL Data Library, May 1995  
A:Description: The complete sequence of a 40.7 kb segment located on the left arm of yeast  
A:Reference number: S55159  
A:Accession: S55181  
A:Molecule type: DNA  
A:Residues: 1-336 <KAT>  
A:Cross-References: EMBL:X87371; NID:g854542; PIDN:CAA60824.1; PID:g854565  
R:Souciat, J.L.; Nagy, M.; le Gouar, M.; Lacroute, F.; Potier, S.  
Gene 79, 59-70, 1989  
A:Title: Organization of the yeast URA2 gene: identification of a defective dihydroorotase  
A:Reference number: S05766; MUID:89378778; PMID:2570735  
A:Accession: S05768  
A:Molecule type: DNA  
A:Residues: 1-124 <SOU>  
A:Cross-References: EMBL:M27174  
R:Katsoulou, C.; Tzermia, M.; Alexandraki, D.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56912  
A:Accession: S56913  
A:Molecule type: DNA

A:Residues: 1-356 <KAW>  
A:Cross-References: EMBL:Z49406; NID:g1008333; PIDN:CAA89426.1; PID:g1008334; MIPS  
R:Katsoulou, C.; Tzermia, M.; Tavernarakis, N.; Alexandraki, D.  
Yeast 12, 787-797, 1996  
A:Title: Sequence analysis of a 40.7 kb segment from the left arm of yeast chromosome  
of chromosome XI.  
A:Reference number: S71643; MUID:96408771; PMID:8813765  
A:Accession: S71666  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-356 <KAF>  
A:Cross-References: EMBL:X87371; NID:g854542; PIDN:CAA60824.1; PID:g854565  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995  
C:Genetics:  
A:Cross-References: SGD:S0003667  
A:Map position: 10L

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 356;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 ELKKER 27  
|||||||  
Db 334 ELKKER 340

RESULT 14  
F82871  
conserved hypothetical membrane lipoprotein UU602 [Imported] - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: F82871  
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of  
A:Reference number: A82870  
A:Accession: F82871  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-383 <GLA>  
A:Cross-References: GB:AE002158; GB:AF222894; NID:g6899599; PIDN:AAF31016.1; GSPDB:  
A:Experimental source: serovar 3; biovar 1  
C:Genetics:  
A:Gene: UU602  
A:Genetic code: SCC3

Query Match  
Best Local Similarity 7.2%; Score 7; DB 2; Length 383;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 KKKKKK 30  
|||||||  
Db 30 KKKKKK 36

RESULT 15  
C84905  
probable extensin [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: C84905  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: C84905  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-394 <STO>  
A:Cross-References: GB:AE002093; NID:g3831447; PIDN:AAC69930.1; GSPDB:GN00139  
C:Genetics:

S41689  
cytochrome-c oxidase (EC 1.9.3.1) chain I - Theileria parva mitochondrion (fragment)  
C:Species: mitochondrion Theileria parva  
C>Date: 31-Dec-1993 #sequence\_revision 02-Aug-1994 #text\_change 07-Dec-1999  
C:Accession: S41689; S40154  
R:Kairo, A.; Fairlamb, A.H.; Gobright, E.; Nene, V.  
EMBO J. 13, 898-905, 1994  
A:Title: A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA sequence  
A:Reference number: S41689; MUID:94155854; PMID:8112303  
A:Accession: S41689  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-483 <KAI>  
A:Cross-references: EMBL:Z23263; NID:9437862; PIDN:CAA80798.1; PID:9437863  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1993  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SCC6  
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane  
transmembrane protein  
F:18-466/Domain: cytochrome-c oxidase chain I homology <Col>  
F:71.386/Binding site: heme a iron (His) (axial ligands) #status predicted  
F:248-252/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F:252/Binding site: oxygen (Tyr) #status predicted  
F:384/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 7.2%; Score 7; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 GIDFIIIF 46  
|||||  
DB 150 GIDFIIIF 156

RESULT 19  
F71619  
hypothetical protein PFB0235w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C>Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: F71619  
R:Gardner, M.J.; Saitzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith,  
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith,  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: F71619  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-483 <GAR>  
A:Cross-references: GB:AE001382; GB:AE001362; NID:93845130; PIDN:AACT1836.1; PID:93  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0235w

Query Match 7.2%; Score 7; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKKK 30  
|||||  
DB 459 KKERKKK 465

RESULT 20  
T04465  
hypothetical protein F4D11.190 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C:Accession: T04465  
R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Hoheisel, J.; Mewer;  
submitted to the Protein Sequence Database, April 1998

A:Gene: At2g46630  
A:Map position: 2

Query Match 7.2%; Score 7; DB 2; Length 394;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 PPSPKVEV 72  
|||||  
DB 124 PPSPKVEV 130

RESULT 16  
AF3271  
probable allantoin permease [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: AF3271  
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
R.; DelVecchio, V.G.; Gotsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
.; Mazur, M.; Voigt, U.S.A. 99, 443-448, 2002  
Proc. Natl. Acad. Sci. U.S.A.  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AF3271  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-409 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AAL51337.1; PID:917982035; GSPDB:GN00190  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI0155  
A:Map position: 1

Query Match 7.2%; Score 7; DB 2; Length 409;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SOTELRK 24  
|||||  
DB 396 SOTELRK 402

RESULT 17  
AH2025  
hypothetical protein all1758 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AH2025  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH2025  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-463 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAF73457.1; PID:gl17130848; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all1758

Query Match 7.2%; Score 7; DB 2; Length 463;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKKER 27  
|||||  
DB 199 ELRKKER 205

RESULT 18

A:Reference number: 215360  
 A:Accession: T04465  
 A:Molecule type: DNA  
 A:Residues: 1-557 <BEV>  
 A:Cross-references: EMBL:AL022537  
 A:Experimental source: cultivar Columbia; BAC clone F4D11  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 107/3; 196/1; 315/2; 388/3; 497/2  
 A:Note: F4D11.190

Query Match 7.2%; Score 7; DB 2; Length 557;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30  
 |||||  
 DB 163 KKERKK 169

## RESULT 21

F70719  
 hypothetical protein RV0976c - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: F70719  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.M.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: F70719  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-560 <COL>  
 A:Cross-references: GB:Z79700; GB:AL123456; NID:g3261628; PIDN:CAB02013.1; PID:ei300054;  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV0976c

Query Match 7.2%; Score 7; DB 2; Length 560;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 REMLTGG 81  
 |||||  
 DB 18 REMLTGG 24

## RESULT 22

S46265  
 cysteine proteinase - Plasmodium vivax  
 C:Species: Plasmodium vivax  
 C:Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 29-Sep-1999  
 C:Accession: S46265  
 R:Rosenthal, P.J.; King, C.S.; Chen, X.; Cohen, F.E.  
 J. Mol. Biol. 241, 312-316, 1994  
 A:Title: Characterization of a Plasmodium vivax cysteine proteinase gene identifies unique  
 A:Reference number: S46265; MUID:94334995; PMID:8057374  
 A:Accession: S46265  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-583 <ROS>  
 A:Cross-references: EMBL:L26362; NID:g431433; PIDN:AAA60368.1; PID:g431434  
 C:Superfamily: trophozoite cysteine proteinase

Query Match 7.2%; Score 7; DB 2; Length 583;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKRE 32

DB 175 ERKKRE 181  
 |||||

## RESULT 23

T01467  
 hypothetical protein T24H24.5 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 04-Mar-2000  
 C:Accession: T01467  
 R:Courtney, L.; Stoneking, T.; Langston, Y.; Mead, K.  
 submitted to the EMBL Data Library, August 1998  
 A:Description: The sequence of A. thaliana T24H24.  
 A:Reference number: Z14333  
 A:Accession: T01467  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-628 <COU>  
 A:Cross-references: EMBL:AF075598; NID:g3293581; PID:g3377830  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 53/2; 216/1; 289/1; 363/3; 413/3  
 A:Note: T24H24.5  
 C:Superfamily: Arabidopsis thaliana hypothetical protein T24H24.5

Query Match 7.2%; Score 7; DB 2; Length 628;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTEL 22  
 |||||  
 DB 222 TLSQTEL 228

## RESULT 24

C69423  
 DNA helicase homolog - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
 C:Accession: C69423  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Do-  
 gleok, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Overbeek, R.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Do-  
 gleok, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch-  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: C69423  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-648 <KLE>  
 A:Cross-references: GB:AE001009; GB:AE000782; NID:g2689332; PIDN:AAB89860.1; PID:g268-  
 C:Superfamily: probable DNA helicase MJ0104

Query Match 7.2%; Score 7; DB 2; Length 648;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 GDSETLS 18  
 |||||  
 DB 609 GDSETLS 615

## RESULT 25

C86333  
 hypothetical protein T20H2.23 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
 C:Accession: C86333  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

Tue May 13 12:12:44 2003

```

ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Marti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86333
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <SPD>
A:Cross-references: GB:AE005172; NID:g8778999; PIDN:AAF79914.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match          7.2%  Score 7;  DB 2;  Length 651;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 28 KKKRERK 34
      |||||
Db 507 KKKRERK 513

RESULT 26
G75638
modulation protein-related protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: G75638
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Shen, M.; Vanathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
Science 286, 1571-1577, 1999
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: G75638
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-703 <WHI>
A:Cross-references: GB:AE001827; NID:g6460959; PIDN:AAF12674.1; PID:g6460971; TIGR:DRC00
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRC0037
A:Map position: plasmid
A:Genome: plasmid
A:Note: plasmid CP1

Query Match          7.2%  Score 7;  DB 2;  Length 703;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 GDSFELS 18
      |||||
Db 165 GDSFELS 171

RESULT 27
T04768
hypothetical protein T16H5.230 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C:Accession: T04768
R:Bevan, M.; De Haan, M.; Maarse, A.C.; Grivell, L.A.; Bancroft, I.; Mewes, H.W.; Mayer,
submitted to the Protein Sequence Database, June 1998
A:Reference number: Z15383
A:Accession: T04768
A:Molecule type: DNA
A:Residues: 1-777 <BEV>
A:Cross-references: EMBL:AL024486
A:Experimental source: cultivar Columbia; BAC clone T16H5

C:Genetics:
A:Map position: 4
A:Introns: 371/1
A:Note: T16H5.230

Query Match          7.2%  Score 7;  DB 2;  Length 777;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 ATRSHLG 92
      |||||
Db 455 ATRSHLG 461

RESULT 28
A46613
protein 4.1, p4.1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 25-Aug-1995
C:Accession: A46613
R:Huang, J.P.; Tang, C.J.; Kou, G.H.; Marchesi, V.T.; Benz Jr., E.J.; Tang, T.K.
J. Biol. Chem. 268, 3758-3766, 1993
A:Title: Genomic structure of the locus encoding protein 4.1. Structural basis for c
C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology
F:213-488/Domain: protein 4.1 membrane-binding domain homology <B41>
A:Reference number: A46613; MUID:93155238; PMID:8429050
A:Accession: A46613
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-858 <HUA>
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBIN:124466, NCBIP:124467)
C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology
F:213-488/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match          7.2%  Score 7;  DB 2;  Length 858;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKKRER 33
      |||||
Db 641 RKKKRER 647

RESULT 29
T28937
hypothetical protein C52B9.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T28937
R:Nelson, J.
submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid C52B9.
A:Reference number: Z20545
A:Accession: T28937
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1257 <NEL>
A:Cross-references: EMBL:U64598; PIDN:AA47974.1; GSPDB:GN00028; CESP:C52B9.8
A:Experimental source: strain Bristol N2; clone C52B9
C:Genetics:
A:Gene: CESP:C52B9.8
A:Map position: X
A:Introns: 15/2; 321/3; 450/3; 596/2; 776/2; 823/2; 871/3

Query Match          7.2%  Score 7;  DB 2;  Length 1257;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 QTELRRK 25
      |||||
Db 55 QTELRRK 61

RESULT 30

```

T16507  
 hypothetical protein F59A6.5 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T16507  
 R:Nhan, M.  
 submitted to the EMBL Data Library, December 1995  
 A:Description: The sequence of C. elegans cosmid F59A6.  
 A:Reference number: Z18526  
 A:Accession: T16507  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1286 <NHA>  
 A:Cross-references: EMBL:U41994; NID:g1123047; PID:g1123050; PIDN:AAA83454.1; CESP:F59A6  
 C:Genetics:  
 A:Gene: CESP:F59A6.5  
 A:Introns: 35/3; 335/3; 695/3; 973/3; 1097/3

Query Match  
 Best Local Similarity 7.2%; Score 7; DB 2; Length 1286;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ASLGDSE 15  
 |||||

Db 317 ASLGDSE 323

RESULT 31  
 S69632  
 regulatory protein SPP41 - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein YDR464W  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 23-Mar-2001  
 C:Accession: S69632; S47864  
 R:Dietrich, F.S.  
 submitted to the EMBL Data Library, August 1995  
 A:Description: The sequence of S. cerevisiae cosmids 9410, 8035, 8166, and 9787.  
 A:Reference number: S69554  
 A:Accession: S69632  
 A:Molecule type: DNA  
 A:Residues: 1-1435 <DIE>  
 A:Cross-references: EMBL:U33050; NID:g927726; PID:g927729; MIPS:YDR464W  
 R:Maddock, J.R.; Weidenhammer, E.M.; Adams, C.C.; Lunz, R.L.; Woolford Jr., J.L.  
 A:Title: Extragenic suppressors of Saccharomyces cerevisiae prp4 mutations identify a new  
 A:Reference number: S47864; MUID:94274035; PMID:8005438  
 A:Accession: S47864  
 A:Molecule type: DNA  
 A:Residues: 1-1394, 'R', <MAD>  
 A:Cross-references: EMBL:U03673; NID:g435022; PIDN:AAA20494.1; PID:g435023  
 C:Genetics:  
 A:Gene: SGD:SPP41  
 A:Cross-references: SGD:S0002872; MIPS:YDR464W  
 A:Map position: 4R  
 C:Keywords: nucleus

Query Match  
 Best Local Similarity 7.2%; Score 7; DB 2; Length 1435;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKKK 30  
 |||||

Db 695 KKERKKK 701

RESULT 32  
 F86342  
 F9H16.4 protein - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Nov-2001  
 C:Accession: F86342  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Ma  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: F86342  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1551 <STO>  
 A:Cross-references: GB:AE0051172; NID:g4836891; PIDN:AAD30594.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match  
 Best Local Similarity 7.2%; Score 7; DB 2; Length 1551;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKRE 32  
 |||||

Db 1304 ERKKRE 1310

RESULT 33  
 T18402  
 asparagine/aspartate rich protein - malaria parasite (Plasmodium falciparum) (fragm  
 C:Species: Plasmodium falciparum  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
 C:Accession: T18402  
 R:Barale, J.C.; Candelle, D.; Attal-Bonnefoy, G.; Dehoux, P.; Bonnefoy, S.; Ridley,  
 Infect. Immun. 65, 3003-3010, 1997  
 A:Title: Plasmodium falciparum AAP1, a giant protein containing repeated motifs ri  
 A:Reference number: Z18929; MUID:97378065; PMID:9234746  
 A:Accession: T18402  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3844 <BAR>  
 A:Cross-references: EMBL:Y08926; NID:el154302; PID:el11435; PIDN:CAA70130.1  
 C:Genetics:  
 A:Gene: aarp1

Query Match  
 Best Local Similarity 7.2%; Score 7; DB 2; Length 3844;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKKK 30  
 |||||

Db 3134 KKERKKK 3140

RESULT 34  
 D69398  
 hypothetical protein AF1189 - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: D69398  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Do  
 Glodek, A.; Zhou, L.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.  
 Nature 390, 364-370, 1997  
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: D69398  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-63 <KLE>  
 A:Cross-references: GB:AE001022; GB:AE000782; NID:g2689345; PIDN:AAB90067.1; PID:g264

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: B84145  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans*  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: B84145  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-89 <STO>  
 A:Cross-references: GB:AP001520; GB:BA000004; NID:g10176401; PIDN:BA807681.1; GSPDB:  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH3962

Query Match 6.2%; Score 6; DB 2; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KKKRR 33  
 Db 84 KKKRR 89

RESULT 38

T04102

smt3 protein - rice

C:Species: *Oryza sativa* (rice)  
 C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000

C:Accession: T04102

R:Leopenta, V.; Chirazzi, P.; Vanderspek, P.; Pizzuti, A.; Hanaoka, F.; Brahe, C.

Genomics 40, 362-366, 1997

A:Title: Smt3a, a human homolog of the *s-cerevisiae*-smt3 gene, maps to chromosome-2

A:Reference number: 205239; MUID:97237059; PMID:9119407

A:Accession: T04102

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-100 <LAP>

A:Cross-references: EMBL:X99608; NID:g1668772; PIDN:CAA67922.1; PID:g1668773

C:Genetics:

A:Gene: smt3

C:Superfamily: yeast Smt3 protein

Query Match 6.2%; Score 6; DB 2; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TGGCLP 84  
 Db 94 TGGCLP 99

RESULT 39

T17566

hypothetical protein a76L - *Chlorella* virus PBCV-1

C:Species: *Chlorella* virus PBCV-1

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T17566

R:Graves, M.V.; Van Etten, J.L.

submitted to the EMBL Data Library, May 1999

A:Reference number: 218806

A:Accession: T17566

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-103 <GRA>

A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96444.1

A:Experimental source: specific host *Chlorella* strain NC64A

C:Genetics:

A:Gene: a76L

C:Superfamily: *Chlorella* virus PBCV-1 hypothetical protein a76L

Query Match 6.2%; Score 6; DB 2; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 66;

Query Match 6.2%; Score 6; DB 2; Length 63;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKR 31  
 Db 51 ERKKR 56

RESULT 35

AG3625

hypothetical protein BMEII0928 [imported] - *Brucella melitensis* (strain 16M)

C:Species: *Brucella melitensis*

C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C:Accession: AG3625

R:DelVecchio, V.G.; Kapral, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,

; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*

A:Reference number: AB3252; PMID:11756688

A:Accession: AG3625

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-74 <KUR>

A:Cross-references: GB:AE008918; PIDN:AAL54170.1; PID:g17985136; GSPDB:GN00191

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEII0928

A:Map position: II

Query Match 6.2%; Score 6; DB 2; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 LLFSHH 57  
 Db 63 LLFSHH 68

RESULT 36

G84773

hypothetical protein At2g35850 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: G84773

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: G84773

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-80 <STO>

A:Cross-references: GB:AE002093; NID:g4510384; PIDN:AAD21472.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g35850

A:Map position: 2

Query Match 6.2%; Score 6; DB 2; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKK 30  
 Db 41 KERKK 46

RESULT 37

B84145

hypothetical protein BH3962 [imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KKRERK 34  
 |||||  
 Db 95 KKRERK 100

## RESULT 40

T07618

cold stress protein cill.8 - garden pea

C:Species: Pisum sativum (garden pea)

C&gt;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Nov-1999

C:Accession: T07618

R:Kung, C.C.; Yeh, K.W.; Lin, C.Y.; Chen, Y.M.

Bot. Bull. Acad. Sin. 39, 9-15, 1998

A:Title: Characterization of a pea gene responsive to low temperature.

A:Reference number: Z16055

A:Accession: T07618

A&gt;Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-110 &lt;KUN&gt;

A:Cross-references: EMBL:U24398; NID:g2947080; PID:g2947081

A:Experimental source: cv. Taichung 9

C:Comment: this protein is cold-induced.

C:Superfamily: cold stress protein COR19

## Query Match

Best Local Similarity 6.2%; Score 6; DB 2; Length 110;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKRERK 29

|||||

Db 88 KKRERK 93

## RESULT 41

B40617

transcription repressor of ansAB operon ansR - Bacillus subtilis

C:Species: Bacillus subtilis

C&gt;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000

C:Accession: B40617; B69586

R:Sun, D.; Setlow, P.

J. Bacteriol. 175, 2501-2506, 1993

A:Title: Cloning and nucleotide sequence of the Bacillus subtilis ansR gene, which encod

A:Reference number: A40617; MUID:93239674; PMID:8478318

A:Accession: B40617

A&gt;Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-116 &lt;SUN&gt;

A:Cross-references: GB:L08205; NID:g304138; PID:AAAT2333.1; PID:g387575

A:Note: sequence extracted from NCBI backbone (NCBI:130324, NCBI:130334)

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Barten

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd

A.; Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hoisappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: B69586

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-116 &lt;KUN&gt;

A:Cross-references: GB:Z99116; GB:AL009126; NID:g2634723; PID:CAB14291.1; PID:g2634794

A:Experimental source: strain 168

## C:Genetics:

A:Gene: ansR

C:Superfamily: probable transcription repressor yowR

C:Keywords: DNA binding; transcription regulation

## Query Match

Best Local Similarity 6.2%; Score 6; DB 2; Length 116;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKK 25

|||||

Db 7 TELRKK 12

## RESULT 42

H72386

mazG-related protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C&gt;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: H72386

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genom

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: H72386

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-116 &lt;ARN&gt;

A:Cross-references: GB:AE001716; GB:AE000512; NID:g4980853; PIDN:AAD35447.1; PID:g49

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0360

## Query Match

Best Local Similarity 6.2%; Score 6; DB 2; Length 116;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKK 30

|||||

Db 110 KERKK 115

## RESULT 43

D72336

transcription regulator, GntR family - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C&gt;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: D72336

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: D72336

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-121 &lt;ARN&gt;

A:Cross-references: GB:AE001746; GB:AE000512; NID:g4981285; PIDN:AAD35848.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0766

C:Superfamily: transcription regulator gntR-related protein ytrA

## Query Match

Best Local Similarity 6.2%; Score 6; DB 2; Length 121;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELRKK 25

|||||

Db 90 TELRKK 95



Tue May 13 12:12:44 2003

## RESULT 44

F84868  
 Hypothetical protein At2g43640 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: F84868  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Lin, X.; Kaul, S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;  
 M.; Koo, H.; Moffat, K.S.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: F84868  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-121 <STO>  
 A:Cross-references: GB:AE002093; NID:g2281106; PIDN:AB64042.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g43640  
 A:Map position: 2

Query Match 6.2%; Score 6; DB 2; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KKRERK 34  
 |||||  
 Db 94 KKRERK 99

## RESULT 45

T37059  
 Hypothetical protein SCJ21.10 - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T37059  
 R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, August 1999  
 A:Reference number: Z21620  
 A:Accession: T37059  
 A:Status: preliminary; translated from GB/EMBL/DDBU  
 A:Molecule type: DNA  
 A:Residues: 1-124 <SEE>  
 A:Cross-references: EMBL:AL109747; PIDN:CAE52356.1; GSPDB:GN00070; SCOEDB:SCJ21.10  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCOEDB:SCJ21.10

Query Match 6.2%; Score 6; DB 2; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TGGCLP 84  
 |||||  
 Db 55 TGGCLP 60

Search completed: May 11, 2003, 20:13:40  
 Job time : 34.0443 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 20:01:37 ; Search time 19.7434 seconds

(without alignments)  
203.775 Million cell updates/sec

Title: US-09-854-133-586

Perfect score: 97

Sequence: 1 EVESRDHASLGDSETLSQT.....LTGCLPWATKSHLGRKCS 97

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	8.2	67	1 RS21_AQUAE	O67028 aquifex aeo
2	7	7.2	129	1 RS9_CHLTR	O84128 chlamydia t
3	7	7.2	133	1 RS9_CHLMU	O9pkx2 chlamydia m
4	7	7.2	134	1 RS9_CHLPN	O928t8 chlamydia p
5	7	7.2	216	1 RGSJ_MOUSE	O9cx84 mus musculus
6	7	7.2	310	1 YDHB_ECOLI	P37598 escherichia
7	7	7.2	356	1 YJN1_YEAST	P47015 saccharomyc
8	7	7.2	440	1 SOCA_HUMAN	Q8wxh5 homo sapien
9	7	7.2	583	1 CYSF_PLAVI	P42666 plasmodium
10	7	7.2	802	1 Y250_HUMAN	Q92540 homo sapien
11	7	7.2	858	1 41_MOUSE	P48193 mus musculus
12	7	7.2	864	1 41_HUMAN	P11171 homo sapien
13	7	7.2	1395	1 SP41_YEAST	P38904 saccharomyc
14	6	6.2	100	1 SMT3_ORYSA	P55857 oryza sativ
15	6	6.2	116	1 ANSR_BACSU	O07683 bacillus su
16	6	6.2	121	1 SR14_ARATH	O04421 arabidopsis
17	6	6.2	129	1 RBFA_PSEAE	O9hv56 pseudomonas
18	6	6.2	132	1 RBFA_VIBCH	O9ku79 vibrio chol
19	6	6.2	132	1 SR14_ORYSA	O04433 oryza sativ
20	6	6.2	137	1 RR9_FORPU	P51291 porphyra pu
21	6	6.2	162	1 Y157_AQUAE	O66547 aquifex aeo
22	6	6.2	166	1 BAIE_EUBSP	P19412 eubacterium
23	6	6.2	184	1 TD52_HUMAN	P55327 homo sapien
24	6	6.2	184	1 TD52_RABIT	O95212 oryctolagus
25	6	6.2	185	1 TD52_MOUSE	O62393 mus musculus
26	6	6.2	188	1 P5PC_RABIT	P22398 oryctolagus
27	6	6.2	211	1 RCSA_ERWAM	P20098 erwinia any
28	6	6.2	211	1 RCSA_ERWST	P27488 erwinia ste
29	6	6.2	231	1 NTA2_MALZE	P39871 zea mays (m
30	6	6.2	233	1 MTMU_MYCSP	P43641 mycoplasma
31	6	6.2	234	1 A29B_DROSI	Q9U968 drosophila
32	6	6.2	237	1 BIOD_YEAST	P53630 saccharomyc
33	6	6.2	238	1 CW14_YEAST	O13547 saccharomyc

34	6	6.2	249	1 YCND_BACSU	P94424 bacillus su
35	6	6.2	252	1 AGL6_ARATH	P29386 arabidopsis
36	6	6.2	256	1 YQEU_BACSU	P54461 bacillus su
37	6	6.2	269	1 FPG_VIBCH	Q9kvc5 vibrio chol
38	6	6.2	279	1 NRTB_PHOLA	Q51881 phormidium
39	6	6.2	283	1 NADC_METJA	Q57916 methanococc
40	6	6.2	286	1 PARB_CHLPN	Q827m0 chlamydia p
41	6	6.2	313	1 EFTS_ANASP	Q8ymy3 anabaena sp
42	6	6.2	315	1 CPPL_ENTHI	Q01957 entamoeba h
43	6	6.2	316	1 L767_CAEEL	Q09517 caenorhabdi
44	6	6.2	319	1 H963_HUMAN	O14626 homo sapien
45	6	6.2	319	1 PRIM_BUCAP	P32000 buchnera ap
46	6	6.2	322	1 YMX7_CAEEL	P34515 caenorhabdi
47	6	6.2	339	1 RPOA_AEGTA	P92429 aegilops ta
48	6	6.2	339	1 RPOA_AGRCR	P92209 agropyron c
49	6	6.2	339	1 RPOA_AUSVE	P33993 australopyr
50	6	6.2	339	1 RPOA_BROIN	P92220 bromus iner
51	6	6.2	339	1 RPOA_CRIDE	P92225 crithopsis
52	6	6.2	339	1 RPOA_EREDI	P93974 eremopyrum
53	6	6.2	339	1 RPOA_FESFE	P93968 festucopsis
54	6	6.2	339	1 RPOA_FESSE	P93956 festucopsis
55	6	6.2	339	1 RPOA_HORVU	P92392 hordeum vul
56	6	6.2	339	1 RPOA_PSAFR	P92418 psathyrosta
57	6	6.2	339	1 RPOA_PSAFU	P93960 psathyrosta
58	6	6.2	339	1 RPOA_PSAST	P93962 psathyrosta
59	6	6.2	339	1 RPOA_SECT	P93964 secale stri
60	6	6.2	339	1 RPOA_THIBE	P92439 thinopyrum
61	6	6.2	339	1 RPOA_WHEAT	P12073 triticum ae
62	6	6.2	341	1 COA2_POVMK	P45966 mouse polyo
63	6	6.2	351	1 ADH_CLOBE	P25984 clostridium
64	6	6.2	357	1 GBA2_DICDI	P16051 dictyosteli
65	6	6.2	359	1 COBT_ECOLI	P36562 escherichia

# ALIGNMENTS

RESULT 1				
RS21_AQUAE	STANDARD;	PRT;	67 AA.	
ID	AC	O67028;		
DT	30-MAY-2000	(Rel. 39, Created)		
DT	30-MAY-2000	(Rel. 39, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	30S ribosomal protein S21.			
GN	RPSU OR AQ_867 OR AQ_867A.			
OS	Aquifex aeolicus.			
OC	Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;			
OC	Aquifex.			
OX	NCBI_TaxID=63363;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VF5;			
RX	MEDLINE=98196666; PubMed=9537320;			
RA	Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,			
RA	Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujaay M., Huber R.,			
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;			
RT	"The complete genome of the hyperthermophilic bacterium Aquifex			
RT	aeolicus."			
RL	Nature 392:353-358(1998).			
CC	-1- SIMILARITY: BELONGS TO THE S21P FAMILY OF RIBOSOMAL PROTEINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; A8000711; AAC06990.1; -			
DR	InterPro; IPR001911; Ribosomal_S21.			
DR	Pfam; PF01165; Ribosomal_S21; 1.			

```

DR PRINTS; PR00976; RIBOSOMALS21.
DR PRODOM; PD005521; Ribosomal_S21; 1.
DR TIGRFAMS; TIGR00030; S21P; 1.
DR PROSITE; PS01181; RIBOSOMALS21; FALSE_NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 67 AA; 8261 MW; E1897087A487EF70 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 67;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKRR 33
DB 43 ERKKRR 50

RESULT 2
RS9_CHLTPR
ID RS9_CHLTPR STANDARD; PRT; 129 AA.
AC 084128;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S9.
GN RPSI OR RS9 OR CP126.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/JW-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis";
RL Science 282:754-759(1998).
CC -!- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001286; AAC67717.1;
DR InterPro; IPR000754; Ribosomal_S9.
DR Pfam; PF00380; Ribosomal_S9; 1.
DR PRODOM; PD001627; Ribosomal_S9; 1.
DR PROSITE; PS00360; RIBOSOMAL_S9; FALSE_NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 129 AA; 14560 MW; 8AF05C6A824D2F8D CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 129;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKKRK 29
DB 108 RKKKRK 114

RESULT 3
RS9_CHLMU
ID RS9_CHLMU STANDARD; PRT; 133 AA.
AC 09PKR2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S9.
GN RPSI OR RS9 OR CPN0246 OR CP0516.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
DR EMBL; AE002307; AAF39259.1;
DR TIGR; TC0402;
DR InterPro; IPR000754; Ribosomal_S9.
DR Pfam; PF00380; Ribosomal_S9; 1.
DR PRODOM; PD001627; Ribosomal_S9; 1.
DR PROSITE; PS00360; RIBOSOMAL_S9; FALSE_NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 133 AA; 14992 MW; DOE7BA313CBAAF35 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 133;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKKRK 29
DB 112 RKKKRK 118

RESULT 4
RS9_CHLPPN
ID RS9_CHLPPN STANDARD; PRT; 134 AA.
AC 09Z8T6; Q9JQJ9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S9.
GN RPSI OR RS9 OR CPN0246 OR CP0516.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
DR EMBL; AE002307; AAF39259.1;
DR TIGR; TC0402;
DR InterPro; IPR000754; Ribosomal_S9.
DR Pfam; PF00380; Ribosomal_S9; 1.
DR PRODOM; PD001627; Ribosomal_S9; 1.
DR PROSITE; PS00360; RIBOSOMAL_S9; FALSE_NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 133 AA; 14992 MW; DOE7BA313CBAAF35 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 133;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKKRK 29
DB 112 RKKKRK 118

```

RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RL from Japan and CWL029 from USA.";  
 CC -1- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE001610; AAD18399.1; -  
 CC EMBL: AE002211; AAF38343.1; -  
 CC EMBL: AP002345; BAA98456.1; -  
 CC TIGR: CP0516; -  
 CC InterPro: IPR000754; Ribosomal\_S9.  
 CC Pfam: PF00380; Ribosomal\_S9; 1.  
 CC ProDom: PD001627; Ribosomal\_S9; 1.  
 CC ProSITE: PS00360; RIBOSOMAL\_S9; 1.  
 CC Ribosomal protein; Complete proteome.  
 KW SEQUENCE 134 AA; 15195 MW; A9869207154FAE8 CRC64;  
 SQ  
 Query Match 7.2%; Score 7; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 RKKERKK 29  
 Db 113 RKKERKK 119  
 RESULT 5  
 RGSJ\_MOUSE STANDARD; PRT; 216 AA.  
 ID RGSJ\_MOUSE  
 AC Q9CX84; Q99L50;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Regulator of G-protein signaling 19 (RGS19).  
 GN RGS19.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gofobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo I., Nikaudo I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast tumor;  
 RA Strausberg R.;  
 RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE  
 CC ACTIVITY OF G PROTEIN ALPHA SUBUNIT THEREBY DRIVING THEM INTO  
 CC THEIR INACTIVE GDP-BOUND FORM. BINDS TO G-ALPHA SUBFAMILY 1  
 CC MEMBERS, WITH THE ORDER G(1)A3 > G(1)A1 > G(O)A >> G(2)A/G(1)A2  
 CC ACTIVITY ON G(2)-ALPHA IS INHIBITED BY PHOSPHORYLATION AND  
 CC PALMITOYLATION OF THE G-PROTEIN (BY SIMILARITY).  
 CC -1- SUBUNIT: INTERACTS WITH GIPC PD2 DOMAIN (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (By similarity).  
 CC -1- PTM: FATTY ACYLATED. HEAVILY PALMITOYLATED IN THE CYSTEINE STRING  
 CC MOTIF (BY SIMILARITY).  
 CC -1- PM: PHOSPHORYLATED, MAINLY ON SERINE RESIDUES (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 1 RGS DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AK019401; BAB31703.1; -  
 CC EMBL: BC003838; AAH03838.1; -  
 CC HSSP: P49795; 1CMZ.  
 CC MGD: MGI:1915153; Rgs19.  
 CC InterPro: IPR000342; RegL\_Gprotein.  
 CC Pfam: PF00615; RGS; 1.  
 CC PRINTS: PR01301; RGS-PROTEIN.  
 CC ProDom: PD001580; Reg\_of\_prg; 1.  
 CC SMART: SM00315; RGS; 1.  
 CC PROSITE: PSS0132; RGS; 1.  
 CC Signal transduction inhibitor; Membrane; Lipoprotein; Palmitate;  
 KW Phosphorylation.  
 FT DOMAIN 39 49 POLY-CYS.  
 FT DOMAIN 90 206 RGS.  
 FT DOMAIN 207 216 INTERACTS WITH GIPC (BY SIMILARITY).  
 FT MOD\_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 151 151 PHOSPHORYLATION (BY MAPK1 AND MAPK3) (BY  
 FT SIMILARITY).  
 FT CONFLICT 82 82 K -> E (IN REF. 2).  
 FT SEQUENCE 216 AA; 24677 MW; 4F166A6607184F31 CRC64;  
 SQ  
 Query Match 7.2%; Score 7; DB 1; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 66 PPSPEV 72  
 Db 78 PPSPEV 84  
 RESULT 6  
 YDHB\_ECOLI STANDARD; PRT; 310 AA.  
 ID YDHB\_ECOLI  
 AC P37598; P77677;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical transcriptional regulator ydHB.  
 GN YDHB OR B1659 OR Z2682 OR ECS2368.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OX Escherichia.  
 RN NCBI\_TaxID=562, 83334;  
 RP [1]  
 RC SEQUENCE FROM N.A.  
 RA STRAIN-K12 / RE28;  
 RL Eberhardt S.M.R., Richter G., Gimbel W., Werner T., Bacher A.;  
 RN Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RC SEQUENCE FROM N.A.  
 RA STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nasimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map";  
 RL DNA Res. 3:363-377(1996).  
 RN [4]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [5]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12";  
 RL DNA Res. 8:11-22(2001).  
 RN [6]  
 RC SEQUENCE OF 106-310 FROM N.A.  
 RP STRAIN-K12;  
 RX MEDLINE=90126847; PubMed=2404765;  
 RA Meng L.M., Kilstrup M., Nygaard P.;  
 RT "Autoregulation of purR repressor synthesis and involvement of purR  
 in the regulation of purR, purC, purL, purM and guaB expression in  
 Escherichia coli";  
 RL Eur. J. Biochem. 187:373-379(1990).  
 RN [7]  
 RC IDENTIFICATION.  
 RP Rudd K.E.;  
 RL Unpublished observations (FEB-1994).  
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----  
 CC EMBL; X69109; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL; AE000261; AAC74731.1; -;  
 CC EMBL; D90809; BAA15425.1; -;  
 CC EMBL; D90810; BAA15434.1; -;  
 CC EMBL; AE005389; AAG56648.1; -;  
 CC EMBL; AP002558; BAB35791.1; -;  
 CC EMBL; X51368; -; NOT\_ANNOTATED\_CDS.  
 CC EcoGene; EG12140; ydhB.  
 CC InterPro; IPR000847; HTH\_LysR.  
 CC InterPro; IPR005119; LysR\_subst.  
 CC Pfam; PF00126; HTH\_1; 1.  
 CC Pfam; PF03466; LysR\_substrate; 1.  
 CC PRINTS; PR00039; HTHLYSR.  
 CC PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
 KW Hypothetical protein; Transcription regulation; DNA-binding;  
 KW Complete proteome.  
 FT DNA\_BIND 19 38 H-T-H MOTIF (BY SIMILARITY).  
 FT CONFLICT 253 253 K -> E (IN REF. 3).  
 SQ SEQUENCE 310 AA; 35250 MW; BBC7A2F2B2AC351A CRC64;  
 Query Match 7.2%; Score 7; DB 1; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 11 LGDSETL 17  
 Db 288 LGDSETL 294  
 RESULT 7  
 ID YJNL\_YEAST STANDARD; PRT; 356 AA.  
 AC P47015;  
 DC 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Hypothetical 41.5 kDa protein in MRS3-URA2 intergenic region.  
 GN YJL131C OR J0682.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RX STRAIN-S288c / FY1679;  
 RC MEDLINE=96408771; PubMed=8813765;  
 RA Katsoulou C., Tzermia M., Tavernarakis N., Alexandraki D.;  
 RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast  
 chromosome X reveals 14 known genes and 13 new open reading frames  
 including homologues of genes clustered on the right arm of  
 chromosome XI";  
 RL Yeast 12:787-797(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X87371; CAA60824.1; -;  
 CC EMBL; Z49406; CAA89426.1; -;  
 CC SGD; S0003667; YJL131C.  
 KW Hypothetical protein.  
 SQ SEQUENCE 356 AA; 41461 MW; BCF907223417B6C2 CRC64;  
 Query Match 7.2%; Score 7; DB 1; Length 356;

Best Local Similarity 100.0%; Pred. No. 7.8; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELKKER 27  
Db 334 ELKKER 340

RESULT 8  
SOC4\_HUMAN STANDARD; PRT; 440 AA.  
AC Q8WXH5;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Suppressor of cytokine signaling 4 (SOCS-4).  
GN SOCS4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hilson D.J., Alexander W.S., Nicola N.A.; Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
RL -!- FUNCTION: SOCS family proteins form part of a classical negative feedback system that regulates cytokine signal transduction.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SOCS BOX DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC EMBL: AF424815; AAL60517.1; -.  
CC InterPro: IPR000980; SH2.  
CC InterPro: IPR001496; SOCS.  
CC PROSITE: PS50001; SH2; 1.  
CC PROSITE: PS50225; SOCS; 1.  
CC SH2 domain; 286 381  
FT DOMAIN 376 425  
FT DOMAIN 440 AA; 50623 MW; A4A747AB7F6F08C CRC64;  
SQ SEQUENCE 440 AA; 50623 MW; A4A747AB7F6F08C CRC64;

Query Match 7.2%; Score 7; DB 1; Length 440;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LSQTELR 23  
Db 166 LSQTELR 172

RESULT 9  
CYSP\_PLAVI STANDARD; PRT; 583 AA.  
AC P42666;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cysteine proteinase precursor (EC 3.4.22.-).  
OS Plasmodium vivax.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5855;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Salvador I;  
RX MEDLINE=94334995; PubMed=8057374;  
RA Rosenthal P.J., Ring C.S., Chen X., Cohen F.E.;

Characterization of a plasmodium vivax cysteine proteinase gene identifies uniquely conserved amino acids that may mediate the substrate specificity of malarial hemoglobinsas.; J. Mol. Biol. 241:312-316(1994).  
-!- FUNCTION: PROBABLY DEGRADES ERYTHROCYTE HEMOGLOBIN.  
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC EMBL: L26362; AAG60368.1; -.  
CC HSSP: P07858; ICSB.  
CC MEROPS: C01.077; -.  
CC InterPro: IPR000668; Peptidase\_C1.  
CC InterPro: IPR000169; SHprot\_acsite.  
CC Pfam: PF00112; Peptidase\_C1; 1.  
CC PRINTS: PR00705; PAPAIN.  
CC PRODOM: PD000158; Peptidase\_C1; 1.  
CC PROSITE: PS00139; THIOL\_PROTEASE\_CYS; 1.  
CC PROSITE: PS00639; THIOL\_PROTEASE\_HLS; 1.  
CC PROSITE: PS00640; THIOL\_PROTEASE\_ASN; 1.  
CC Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal.  
FT SIGNAL 1 ?  
FT PROPEP 338 ACTIVATION PEPTIDE (POTENTIAL).  
FT CHAIN 339 583 CYSTEINE PROTEINASE.  
FT ACT\_SITE 363 363 BY SIMILARITY.  
FT ACT\_SITE 495 495 BY SIMILARITY.  
FT ACT\_SITE 547 547 BY SIMILARITY.  
FT DISULFID 360 402 BY SIMILARITY.  
FT CARBOHYD 70 70 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 195 195 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 381 381 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 486 486 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 494 494 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 583 AA; 65721 MW; 15E8A840717C406 CRC64;

Query Match 7.2%; Score 7; DB 1; Length 583;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ERKKRE 32  
Db 175 ERKKRE 181

RESULT 10  
Y250\_HUMAN STANDARD; PRT; 802 AA.  
ID Y250\_HUMAN  
AC Q92540;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein KIAA0250.  
GN C10RF16 OR KIAA0250.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=97191544; PubMed=9039502;  
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y., Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;  
RA "Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain.";  
RT

RL DNA Res. 3:321-329(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D87437; BAA13381.1; --  
 DR Genew; HGNC:16792; Clorf16.  
 KW Hypothetical protein.  
 SQ SEQUENCE 802 AA; 88998 MW; E9AFIC0AAL1790023 CRC64;  
  
 Query Match 7.2%; Score 7; DB 1; Length 802;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 18 SQTELRK 24  
 Db 282 SQTELRK 288  
 |||||  
  
 RESULT 11  
 ID 41\_MOUSE STANDARD; PRT; 858 AA.  
 AC P48193;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN Protein 4.1 (Band 4.1) (P4.1) (4.1R).  
 OS EPB41 OR EPB4.1.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN-BALB/c;  
 RX MEDLINE=93155238; PubMed=8429050;  
 RA Huang J.-P., Tang C.-J.C., Kou G.-H., Marchesi V.T., Benz E.J. Jr.,  
 RA Tang T.K.;  
 RT "Genomic structure of the locus encoding protein 4.1. Structural  
 RT basis for complex combinatorial patterns of tissue-specific  
 RT alternative RNA splicing."  
 RL J. Biol. Chem. 268:3758-3766(1993).  
 RN [2]  
 RP SEQUENCE OF 709-713, AND CHARACTERIZATION OF CARBOXY-TERMINAL DOMAIN.  
 RX MEDLINE=21325946; PubMed=11432737;  
 RA Scott C., Phillips G.W., Baines A.J.;  
 RT "Properties of the C-terminal domain of 4.1 proteins."  
 RL Eur. J. Biochem. 268:3709-3717(2001).  
 CC -!- FUNCTION: Protein 4.1 is a major structural element of the  
 CC erythrocyte membrane skeleton. It plays a key role in regulating  
 CC membrane physical properties of mechanical stability and  
 CC deformability by stabilizing spectrin-actin interaction. Binds  
 CC with a high affinity to glycophorin and with lower affinity to  
 CC band III protein.  
 CC -!- ALTERNATIVE PRODUCTS: A number of isoforms are produced by  
 CC alternative splicing.  
 CC -!- MASS SPECTROMETRY: MW=17199.3; METHOD=Electrospray; RANGE=709-858.  
 CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L00919; AAA37123.1; --

DR EMBL; L00919; AAA37122.1; --  
 DR MGD; MGI:95401; Epb4.1.  
 DR InterPro; IPR000299; Band 4.1.  
 DR Pfam; PF00373; Band 41; 1.  
 DR PRINTS; PR00935; BAND41.  
 DR SMART; SM00295; B41; 1.  
 DR PROSITE; PS00660; BAND\_41\_1; 1.  
 DR PROSITE; PS00661; BAND\_41\_2; 1.  
 DR PROSITE; PS00057; BAND\_41\_3; 1.  
 KW Structural protein; Alternative splicing; Cytoskeleton;  
 KW Phosphorylation.  
 FT DOMAIN 208 422 BAND 4.1-LIKE.  
 FT DOMAIN 489 608 HYDROPHILIC.  
 FT DOMAIN 609 707 SPECTRIN--ACTIN-BINDING.  
 FT DOMAIN 710 858 CARBOXYL-TERMINAL (CTD).  
 SQ SEQUENCE 858 AA; 95990 MW; 5P2FEF077946134E CRC64;  
  
 Query Match 7.2%; Score 7; DB 1; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 27 RKKKRR 33  
 Db 641 RKKKRR 647  
 |||||  
  
 RESULT 12  
 ID 41\_HUMAN STANDARD; PRT; 864 AA.  
 AC P11171; P11176; Q9Y578; Q9Y579; Q14245;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein 4.1 (Band 4.1) (P4.1) (EPB4.1) (4.1R).  
 GN EPB41 OR E41P.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (NON-ERYTHROID FORM).  
 RX MEDLINE=88234496; PubMed=3375238;  
 RA Tang T.K., Leto T.L., Correas I., Alonso M.A., Marchesi V.T.,  
 RA Benz E.J. Jr.;  
 RT "Selective expression of an erythroid-specific isoform of protein  
 RT 4.1."  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:3713-3717(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A. (NON-ERYTHROID ISOFORM).  
 RX MEDLINE=89132003; PubMed=3223413;  
 RA Tang T.K., Leto T.L., Marchesi V.T., Benz E.J. Jr.;  
 RT "Expression of specific isoforms of protein 4.1 in erythroid and non-  
 RT erythroid tissues."  
 RL Adv. Exp. Med. Biol. 241:81-95(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 378-393 (ERYTHROID ISOFORM).  
 RC TISSUE-Reticulocytes;  
 RX MEDLINE=87092279; PubMed=3467321;  
 RA Conboy J.G., Kan Y.W., Shohet S.B., Mohandas N.;  
 RT "Molecular cloning of protein 4.1, a major structural element of the  
 RT human erythrocyte membrane skeleton."  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9512-9516(1986).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RX MEDLINE=91217063; PubMed=2022844;  
 RA Conboy J.G., Chan J.Y.C., Chasis J.A., Kan Y.W., Mohandas N.;  
 RT "Tissue- and development-specific alternative RNA splicing regulates  
 RT expression of multiple isoforms of erythroid membrane protein 4.1."  
 RL J. Biol. Chem. 266:8273-8280(1991).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Huang S.C., Wang C., Lichtenauer U., Vortmeyer A., Zhuang Z.;  
 RT "Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5."





CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; 003673; AAA20494.1; -  
 DR PIR; S47864; S47864.  
 DR SGD; S0002872; SPP41.  
 DR InterPro; IPR003503; UTM.  
 KW Nuclear protein.  
 FT DOMAIN 583 699  
 SQ SEQUENCE 1395 AA; 156933 MW; 6EC9DA9BB21B3471 CRC64;

Query Match  
 Best Local Similarity 7.2%; Score 7; DB 1; Length 1395;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 24 KKERKKK 30  
 Db 695 KKERKKK 701  
 |||||

RESULT 14  
 SMT3\_ORYSA  
 ID SMT3\_ORYSA STANDARD; PRT; 100 AA.  
 AC P55857;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE Ubiquitin-like protein SMT3.  
 GN SMT3.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97237059; PubMed=9119407;  
 RA Lapenta V., Chiurazzi P., van der Spek P.J., Pizzuti A.,  
 RA Hanaoka F., Brabe C.;  
 RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to  
 RT chromosome 21qter and defines a novel gene family.";  
 RL Genomics 40:362-367(1997).  
 CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X99608; CAA67922.1; -  
 DR HSPSP; Q93068; IASR.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubi\_qitin; 1.  
 DR SMART; SM00213; Usq; 1.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 FT DOMAIN 19 96  
 SQ SEQUENCE 100 AA; 10928 MW; 624E44BC94C00268 CRC64;

Query Match  
 Best Local Similarity 6.2%; Score 6; DB 1; Length 100;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 79 TGGCLP 84  
 Db 94 TGGCLP 99  
 |||||

RESULT 15  
 ANSR\_BACSU  
 ID ANSR\_BACSU STANDARD; PRT; 116 AA.  
 AC Q07683;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE ANS operon repressor protein.  
 GN ANSR.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93239674; PubMed=8478318;  
 RA Sun D., Setlow P.;  
 RT "Cloning and nucleotide sequence of the Bacillus subtilis ansR gene,  
 RT which encodes a repressor of the ans operon coding for L-asparaginase  
 RT and L-aspartase.";  
 RL J. Bacteriol. 175:2501-2506(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=168 / JH642;  
 RX MEDLINE=97124195; PubMed=8969508;  
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,  
 RA Kobayashi Y.;  
 RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of  
 RT the Bacillus subtilis genome containing the skin element and many  
 RT sporulation genes.";  
 RL Microbiology 142:3103-3111(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Terpstra P., Tognoni A.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Takahashi H., Vassarotti A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Weitzenecker T.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256(1997).  
 CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE ANS OPERON CODING FOR  
 CC L-ASPARAGINASE AND L-ASPARTASE. NH4 MAY INFLUENCE THIS REPRESSION.  
 CC -1- SIMILARITY: BELONGS TO THE PBSX(XRE) FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L08205; AAA72333.1; -  
 DR EMBL: D84432; BAA12641.1; -  
 DR EMBL: Z99116; CAB14291.1; -  
 DR PIR: B40617; B40617.  
 DR Subtilist; BG10299; ansr.  
 DR InterPro: IPR001387; HTH\_3.  
 DR Pfam: PF01381; HTH\_3; 1.  
 DR SMART; SM00530; HTH\_XRE; 1.  
 KW Transcription regulation; Repressor; DNA-binding; Complete proteome.  
 FT DNA\_BIND 17 36 H-T-H MOTIF (BY SIMILARITY).  
 SQ SEQUENCE 116 AA; 13231 MW; 29B86500ACB901F8 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TELKK 25  
 Db 7 TELKK 12

#### RESULT 16

ID SR14\_ARATH STANDARD; PRT; 121 AA.  
 AC 004421; 029839;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Signal recognition particle 14 kDa protein (SRP14).  
 GN SRP14 OR AT2G43640 OR F18019.25.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=cv. Columbia;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-J., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Coppenhaver G.P., Preuss D.,  
 RA Niernan W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,  
 RA Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 402:761-768(1999).  
 CC -!- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE  
 CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC RETICULUM  
 CC MEMBRANE. SRP9 TOGETHER WITH SRP14 AND THE ALU PORTION OF THE SRP  
 CC RNA, CONSTITUTES THE ELONGATION ARREST DOMAIN OF SRP. THE COMPLEX  
 CC OF SRP9 AND SRP14 IS REQUIRED FOR SRP RNA BINDING (BY SIMILARITY).  
 CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE  
 CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,  
 CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE SRP14 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Y10116; CAA71202.1; -  
 DR EMBL: AC002333; AAB64042.1; -  
 DR HSSP: P16254; 1914.  
 DR InterPro: IPR003210; SRP14.  
 DR Pfam: PF02290; SRP14; 1.  
 KW Signal recognition particle; RNA-binding.  
 FT CONFLICT 116 116 P -> T (IN REF. 2).  
 SQ SEQUENCE 121 AA; 13777 MW; 216D2AA83B24E7DD CRC64;

Query Match 6.2%; Score 6; DB 1; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KKRERK 34  
 Db 94 KKRERK 99

#### RESULT 17

ID RBFA\_PSEAE STANDARD; PRT; 129 AA.  
 AC Q9HV36;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribosome-binding factor A.  
 GN RBFA OR PA4743.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=2043737; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -!- FUNCTION: Associates with free 30S ribosomal subunits (but not  
 CC with 30S subunits that are part of 70S ribosomes or polysomes).  
 CC Essential for efficient processing of 16S rRNA. May interact with  
 CC the 5'-terminal helix region of 16S rRNA (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE RBFA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE004888; AAG08129.1; -  
 DR InterPro: IPR000238; Rib\_bind\_factA.  
 DR Pfam: PF02033; RBFA; 1.  
 DR ProDom: PD007327; Rib\_bind\_factA; 1.  
 DR TIGRFAMs: TIGR00082; rbfa; 1.  
 DR PROSITE: PS01319; RBFA; 1.  
 KW rRNA processing; Complete proteome.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

SQ SEQUENCE 129 AA; 14547 MW; 1BE5E7F0C08FA9D6 CRC64;
Query Match 6.2%; Score 6; DB 1; Length 129;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VEVSRD 7
DB 40 VEVSRD 45

RESULT 18
RBFA_VIBCH STANDARD; PRT; 132 AA.
AC Q9KU79;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Ribosome-binding factor A.
GN RBFA OR VC0644.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -!- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5'terminal helix region of 16S rRNA (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE RBFA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE004150; AAF93810.1; AL7_INIT.
CC TIGR; VC0644; -.
CC InterPro; IPR000238; Rib_bind_factA.
CC Pfam; PF02033; RBFA; 1.
CC ProDom; PD007327; Rib_bind_factA; 1.
CC TIGRFAMs; TIGR00082; rbfa; 1.
CC PROSITE; PS01319; RBFA; 1.
CC RNA processing; Complete proteome.
CC SEQUENCE 132 AA; 15404 MW; 26D950866B1C6BA0 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 132;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VEVSRD 7
DB 40 VEVSRD 45

RESULT 19
SR14_ORYSA STANDARD; PRT; 132 AA.
AC Q04433;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Signal recognition particle 14 kDa protein (SRP14).
GN SRP14.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare; TISSUE=Etiolated root;
RA Bui N., Wolff N., Strub K.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE
CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC RETICULUM
CC MEMBRANE. SRP9 TOGETHER WITH SRP14 AND THE ALU PORTION OF THE SRP
CC RNA, CONSTITUTES THE ELONGATION ARREST DOMAIN OF SRP. THE COMPLEX
CC OF SRP9 AND SRP14 IS REQUIRED FOR SRP RNA BINDING (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE
CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,
CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE SRP14 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y10118; CAA71204.1; -.
CC HSSP; P16254; 1914.
CC InterPro; IPR003210; SRP14.
CC Pfam; PF02290; SRP14; 1.
CC Signal recognition particle; RNA-binding.
CC SEQUENCE 132 AA; 14838 MW; 374DA93020890A73 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 132;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KKREK 34
DB 94 KKREK 99

RESULT 20
RR9_PORPU STANDARD; PRT; 137 AA.
AC P51291;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Chloroplast 30S ribosomal protein S9.
GN RPS9.
OS Porphyra purpurea.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport;
RA Reith M.E., Munholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome."
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF RIBOSOMAL PROTEINS.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U38804; AAC08177.1; -;  
DR InterPro; IPR000754; Ribosomal\_S9.  
DR Pfam; PF00380; Ribosomal\_S9; 1.  
DR ProDom; PD001627; Ribosomal\_S9; 1.  
DR PROSITE; PS00360; Ribosomal\_S9; 1.  
KW Ribosomal protein; Chloroplast.  
SQ SEQUENCE 137 AA; 15033 MW; AAGB96DFAE728A57 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 137;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 92 GRKCS 97  
Db 16 GRKCS 21  
|||||

RESULT 21  
Y157\_AQUAE STANDARD; PRT; 162 AA.  
AC 066547;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein AQ\_157 precursor.  
GN AQ\_157.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
OC Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjaj M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus";  
RL Nature 392:353-358(1998).  
CC -1- SIMILARITY: BELONGS TO THE OMPH/HLEPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE000676; AAC06512.1; -;  
KW Hypothetical protein; Coiled coil; Signal; Complete proteome.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 162 HYPOTHETICAL PROTEIN AQ\_157.  
FT DOMAIN 32 118 COILED COIL (POTENTIAL)  
SQ SEQUENCE 162 AA; 18878 MW; 61DB961E197471A6 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 162;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 19 QTELK 24  
Db 43 QTELK 48  
|||||

RESULT 22  
BAIE\_EUBSP STANDARD; PRT; 166 AA.  
ID BAIE\_EUBSP  
AC P19412;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Bile acid-inducible operon protein E.  
DE BAIE.  
GN Eubacterium sp. (strain VPI 12708).  
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=29347;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-31.  
RX MEDLINE=91072253; PubMed=2254270;  
RA Mallonee D.H., White W.B., Hylemon P.B.;  
RT "Cloning and sequencing of a bile acid-inducible operon from  
RT Eubacterium sp. strain VPI 12708";  
RL J. Bacteriol. 172:7011-7019(1990).  
CC -1- PATHWAY: Bile acid catabolism.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U57489; AAC45413.1; -;  
DR PIR; D37844; D37844.  
KW Bile acid catabolism.  
FT CONFLICT 2  
SQ SEQUENCE 166 AA; 19533 MW; 1CBCE86C85ADC3E5 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 68 SPKEVT 73  
Db 63 SPKEVT 68  
|||||

RESULT 23  
TD52\_HUMAN STANDARD; PRT; 184 AA.  
ID TD52\_HUMAN  
AC P55327; Q13056;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tumor protein D52 (N8 protein).  
GN TP052.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Breast carcinoma;  
RX MEDLINE=95316866; PubMed=7796418;  
RX Byrne J.A., Tomasetto C., Garnier J.M., Rouyer N., Mattel M.-G.,  
RX Bellocq J.P., Rio M.C., Basset P.;  
RT "A screening method to identify genes commonly overexpressed in  
RT carcinomas and the identification of a novel complementary DNA  
RT sequence";  
RL Cancer Res. 55:2896-2903(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96197754; PubMed=8632896;

RA Chen S.L., Maroulakou I.G., Green J.E., Romano-Spica V., Modi W.,  
RA Lautenberger J., Bhat N.R.;  
RT "Isolation and characterization of a novel gene expressed in multiple  
RT cancers.";  
RL Oncogene 12:741-751(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Skin;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP INTERACTIONS.  
RX MEDLINE=98143307; PubMed=9484778;  
RA Byrne J.A., Nourse C.R., Basset P., Gunning P.;  
RT "Identification of homo- and heteromeric interactions between members  
RT of the breast carcinoma-associated D52 protein family using the yeast  
RT two-hybrid system.";  
RL Oncogene 16:873-881(1998).  
CC -1- SUBUNIT: FORMS HOMODIMER OR HETERODIMER WITH OTHER MEMBERS OF THE  
CC FAMILY.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN CANCER CELLS.  
CC -1- SIMILARITY: BELONGS TO THE TPD52 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U18914; AAC50183.1; -;  
DR EMBL; S82081; AAB36475.1; -;  
DR EMBL; BC018117; AAI18117.1; -;  
DR Genew; HGNC:12005; TPD52.  
DR MIM; 604068; -;  
KW Coiled coil.  
FT DOMAIN 22 74 COILED COIL (POTENTIAL).  
SQ SEQUENCE 184 AA; 19863 MW; 4821EC86D1C3339D CRC64;  
  
Query Match 6.2%; Score 6; DB 1; Length 184;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 14 SETLSQ 19  
Db 100 SETLSQ 105  
-----  
RESULT 24  
ID TPD52\_RABIT STANDARD; PRT; 184 AA.  
AC Q95212;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tumor protein D52 (28 kDa calcium-dependent phosphoprotein) (pp28).  
GN TPD52 OR CSPP28.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE-Gastric fundus;  
RC MEDLINE=96355320; PubMed=8702730;  
RA Parente J.A., Goldenring J.R., Petropoulos A.C., Hellman U.,  
RA Chew C.S.;  
RT "Purification, cloning, and expression of a novel, endogenous,  
RT calcium-sensitive, 28-kDa phosphoprotein.";  
RL J. Biol. Chem. 271:20096-20101(1996).  
CC -1- SUBUNIT: FORMS HOMODIMER OR HETERODIMER WITH OTHER MEMBERS OF THE  
CC FAMILY (BY SIMILARITY).

CC -1- PTM: PHOSPHORYLATED IN A CALCIUM/CALMODULIN-DEPENDENT MANNER.  
CC -1- SIMILARITY: BELONGS TO THE TPD52 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U35428; AAC48616.1; -;  
DR Coiled coil; Phosphorylation.  
FT DOMAIN 22 74 COILED COIL (POTENTIAL).  
SQ SEQUENCE 184 AA; 19809 MW; 8CEA3C2CD6AC3DC4 CRC64;  
  
Query Match 6.2%; Score 6; DB 1; Length 184;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 14 SETLSQ 19  
Db 100 SETLSQ 105  
-----  
RESULT 25  
ID TPD52\_MOUSE STANDARD; PRT; 185 AA.  
AC Q62393;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tumor protein D52 (md52).  
GN TPD52.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [3]  
RP SEQUENCE FROM N.A.  
RX TISSUE-Mammary gland;  
RC MEDLINE=97001154; PubMed=8812487;  
RA Byrne J.A., Mattei M.-G., Basset P.;  
RT "Definition of the tumor protein D52 (TPD52) gene family through  
RT cloning of D52 homologues in human (hd53) and mouse (md52).";  
RL Genomics 35:523-532(1996).  
CC -1- SUBUNIT: FORMS HOMODIMER OR HETERODIMER WITH OTHER MEMBERS OF THE  
CC FAMILY (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE TPD52 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U44426; AAB40897.1; -;  
DR MGD; MGI:107749; Tpd52.  
KW Coiled coil.  
FT DOMAIN 22 74 COILED COIL (POTENTIAL).  
SQ SEQUENCE 185 AA; 20059 MW; 345B487842135D33 CRC64;  
  
Query Match 6.2%; Score 6; DB 1; Length 185;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 14 SETLSQ 19  
Db 100 SETLSQ 105

RESULT 26  
PSPC\_RABBIT STANDARD; PRT; 188 AA.  
AC P22398;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Pulmonary surfactant-associated protein C precursor (SP-C) (Pulmonary surfactant-associated proteolipid SPL(Va1)).  
GN SFTPC OR SFTP2.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=New Zealand white; TISSUE=Lung;  
RX MEDLINE=92353123; PubMed=1643107;  
RA Connolly I., Possmayer F.;  
RT "CDNA sequence and alternative mRNA splicing of surfactant-associated protein C (SP-C) in rabbit lung.";  
RL Biochim. Biophys. Acta 1127:199-207(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=93118799; PubMed=1335697;  
RA Boggaram V., Margana R.K.;  
RT "Rabbit surfactant protein C: cDNA cloning and regulation of alternatively spliced surfactant protein C mRNAs.";  
RL Am. J. Physiol. 263:L634-L644(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Margana R.K., Boggaram V.;  
RT "Cloning, sequence and characterization of the rabbit surfactant protein C gene.";  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 24-188 FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=93105936; PubMed=1468410;  
RA Durham P.L., Nanthakumar E.J., Snyder J.M.;  
RT "Developmental regulation of surfactant-associated proteins in rabbit fetal lung in vivo.";  
RL Exp. Lung Res. 18:775-793(1992).  
RN [5]  
RP SEQUENCE OF 24-37.  
RX MEDLINE=91200266; PubMed=2015882;  
RA Johansson J., Persson P., Loewenadler B., Robertson B., Joernvall H., Curstedt T.;  
RT "Canine hydrophobic surfactant polypeptide SP-C. A lipopeptide with one thioester-linked palmitoyl group.";  
RL FEBS Lett. 281:119-122(1991).  
CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10% PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X65078; CAA46204.1; ALT\_INIT.  
CC EMBL; S51983; AAB24761.1; -.  
CC EMBL; AF037445; AAC18032.1; -.  
CC EMBL; S51597; AAB24762.1; -.

DR EMBL; S51098; AAB24576.2; -.  
DR PIR; S14815; S14815.  
DR PIR; S19946; S19946.  
DR InterPro; IPR001729; Pulm\_surfact\_AP.  
DR Pfam; PF02125; PSAP; 1.  
DR ProDom; PD009591; Pulm\_surfact\_AP; 1.  
DR SMART; SM00019; SF.P; 1.  
DR PROSITE; PS00341; SURFACT\_PALMITOYL; 1.  
KW Surface film; Gaseous exchange; Lipoprotein; Palmitate.  
FT PROPEP 1 23  
FT CHAIN 24 58 PULMONARY SURFACTANT-ASSOCIATED PROTEIN C.  
FT PROPEP 59 188  
FT LIPID 28 28 PALMITATE.  
FT LIPID 29 29 PALMITATE.  
FT CONFLICT 115 115 P -> PP (IN REF. 2).  
FT CONFLICT 153 153 G -> A (IN REF. 4).  
FT CONFLICT 159 159 A -> G (IN REF. 4).  
FT CONFLICT 161 161 G -> R (IN REF. 1).  
FT CONFLICT 186 186 I -> Y (IN REF. 4).  
SQ SEQUENCE 188 AA; 19836 MW; F622EEA933786F78 CRC64;  
Query Match 6.2%; Score 6; DB 1; Length 188;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 33 RKFOAN 38  
Db 139 RKFOAN 144  
RESULT 27  
RCSA\_ERWAM STANDARD; PRT; 211 AA.  
ID RCSA\_ERWAM  
AC P20098;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-MAR-1992 (Rel. 21, Last annotation update)  
DE Colanic acid capsular biosynthesis activation protein A.  
GN RCSA.  
OS Erwinia amylovora.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Erwinia.  
OX NCBI\_TaxID=552;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=T;  
RX MEDLINE=91132105; PubMed=2283503;  
RA Coleman M., Pearce R., Hitchin E., Busfield F., Mansfield J.W., Roberts I.S.;  
RT "Molecular cloning, expression and nucleotide sequence of the rcsA gene of Erwinia amylovora, encoding a positive regulator of capsule expression: evidence for a family of related capsule activator proteins.";  
RL J. Gen. Microbiol. 136:1799-1806(1990).  
CC -1- FUNCTION: POSITIVE REGULATOR OF CAPSULAR POLYSACCHARIDE SYNTHESIS. RCSA AND RCSB FORM A COMPLEX TO PROMOTE TRANSCRIPTION OF THE GENES FOR CAPSULE SYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M57387; AAA24864.1; -.  
CC InterPro; IPR000792; HTH\_LuxR.  
CC Pfam; PF00196; Gere; 1.  
CC PRINTS; PR00038; HTHLUXR.

```

DR PRODOM; PD000307; HTH_LuxR; 1.
DR SMART; SM00421; HTH_LuxR; 1.
DR PROSITE; PS00622; HTH_LuxR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Activator.
FT DNA_BIND 159 178 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 211 AA; 24349 MW; 51BC162269735C55 CRC64;

Query Match
Best Local Similarity 6.2%; Score 6; DB 1; Length 211;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTE 21
Db 141 TLSQTE 146

RESULT 28
RCSA_ERWST
ID RCSA_ERWST STANDARD; PRT; 211 AA.
AC P27488;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Colanic acid capsular biosynthesis activation protein A.
GN RCSA.
OS Erwinia stewartii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=66271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS104;
RX MEDLINE=91375445; PubMed=1896018;
RA Poetter K., Coplin D.L.;
RT "Structural and functional analysis of the rcsA gene from Erwinia
  stewartii.";
RL Mol. Gen. Genet. 229:155-160(1991).
CC -1- FUNCTION: POSITIVE REGULATOR OF CAPSULAR POLYSACCHARIDE
  SYNTHESIS. RCSA AND RCSB FORM A COMPLEX TO PROMOTE TRANSCRIPTION
  OF THE GENES FOR CAPSULE SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
  REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X58707; CAA41544.1; -
CC PIR; S17701; S17701.
CC InterPro; IPR000792; HTH_LuxR.
CC Pfam; PF00196; GERE; 1.
CC PRINTS; PR00038; HTHLUXR.
CC ProDom; PD000307; HTH_LuxR; 1.
CC SMART; SM00421; HTH_LuxR; 1.
CC PROSITE; PS00622; HTH_LuxR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Activator.
FT DNA_BIND 159 178 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 211 AA; 24238 MW; 5298CE6C21E6AA47 CRC64;

Query Match
Best Local Similarity 6.2%; Score 6; DB 1; Length 211;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTE 21
Db 141 TLSQTE 146

RESULT 29
RCSA_ERWST
ID RCSA_ERWST STANDARD; PRT; 211 AA.
AC P27488;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Colanic acid capsular biosynthesis activation protein A.
GN RCSA.
OS Erwinia stewartii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=66271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS104;
RX MEDLINE=91375445; PubMed=1896018;
RA Poetter K., Coplin D.L.;
RT "Structural and functional analysis of the rcsA gene from Erwinia
  stewartii.";
RL Mol. Gen. Genet. 229:155-160(1991).
CC -1- FUNCTION: POSITIVE REGULATOR OF CAPSULAR POLYSACCHARIDE
  SYNTHESIS. RCSA AND RCSB FORM A COMPLEX TO PROMOTE TRANSCRIPTION
  OF THE GENES FOR CAPSULE SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
  REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X58707; CAA41544.1; -
CC PIR; S17701; S17701.
CC InterPro; IPR000792; HTH_LuxR.
CC Pfam; PF00196; GERE; 1.
CC PRINTS; PR00038; HTHLUXR.
CC ProDom; PD000307; HTH_LuxR; 1.
CC SMART; SM00421; HTH_LuxR; 1.
CC PROSITE; PS00622; HTH_LuxR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Activator.
FT DNA_BIND 159 178 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 211 AA; 24238 MW; 5298CE6C21E6AA47 CRC64;

Query Match
Best Local Similarity 6.2%; Score 6; DB 1; Length 211;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTE 21
Db 141 TLSQTE 146

RESULT 30
MTMU_MYCSP
ID MTMU_MYCSP STANDARD; PRT; 233 AA.
AC P43641;
DT 01-NOV-1995 (Rel. 32, Created)

```

```

NTA2_MAIZE
ID NIA2_MAIZE STANDARD; PRT; 231 AA.
AC P39871;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Nitrate reductase [NAD(P)H] (EC 1.6.6.2) (NR) (Fragment).
GN NAR.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. W64A X W182E; TISSUE=Root;
RA Long D.M., Oaks A., Rothstein S.J.;
RT "Regulation of maize root nitrate reductase mRNA levels.";
RL Physiol. Plantarum 85:561-566(1992).
CC -1- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
  STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -1- CATALYTIC ACTIVITY: NAD(P)H + nitrate -> NAD(P)(+) + nitrite +
  H(2)O.
CC -1- COFACTOR: REQUIRES FAD, A HEME GROUP (CALLED CYTOCHROME B-557)
  AND ONE MOLYBDENUM ATOM.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
  N-TERMINAL DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC -1- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
  C-TERMINAL DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X64446; CAA45776.1; -
CC PIR; S24544; S24544.
CC HSSP; P17571; 2CND.
CC MaizeDB; 25891.
CC InterPro; IPR001199; Cyt_B5.
CC InterPro; IPR001834; Cyt_B5_reductase.
CC InterPro; IPR000572; Euk_Mb_Oxred.
CC InterPro; IPR001433; Oxred_FAD/NAD(P).
CC Pfam; PF00175; NAD_binding_1.
CC Pfam; PF00970; FAD_binding_6; 1.
CC PROSITE; PS00191; CYTOCHROME_B5_1; PARTIAL.
CC PROSITE; PS00255; CYTOCHROME_B5_2; PARTIAL.
CC PROSITE; PS00559; MOLYBDOPTERIN_EUK; PARTIAL.
KW Oxidoreductase; Flavoprotein; FAD; NAD; NADP; Heme; Molybdenum;
  Nitrate assimilation; Multigene family.
FT NON_TER 1
SQ SEQUENCE 231 AA; 26254 MW; 9864B425C0ED45F4 CRC64;

Query Match
Best Local Similarity 6.2%; Score 6; DB 1; Length 231;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSET 16
Db 193 LGDSET 198

RESULT 30
MTMU_MYCSP
ID MTMU_MYCSP STANDARD; PRT; 233 AA.
AC P43641;
DT 01-NOV-1995 (Rel. 32, Created)

```



```

CC -----
DR EMBL; X76192; CAA53786.1; -
DR REBASE; 3449; M.Muni.
DR InterPro; IPR002052; N6_Mtase.
DR PROSITE; PS00092; N6_MTASE; 1.
DR TRANSFERASE; Methyltransferase; Restriction system.
KW SEQUENCE 233 AA; 26933 MW; 0BE21A05BB2A3A2E CRC64;
SQ
Query Match 6.2%; Score 6; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 16 TLSQTE 21
| | | | |
Db 142 TLSQTE 147

RESULT 31
A29B_DROSI
ID A29B_DROSI STANDARD; PRT; 234 AA.
AC Q9U968;
DT 16-OCT-2001 (Rel. 40, Created)
DT DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Accessory gland protein Acp29Ab precursor.
DE ACP29AB.
GN Drosophila simulans (Fruit fly).
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S9;
RX MEDLINE=99282496; PubMed=10353898;
RA Aguade M.;
RT "Positive selection drives the evolution of the Acp29AB accessory
RT gland protein in Drosophila."
RL Genetics 152:543-551(1999).
CC - FUNCTION: RESPONSIBLE FOR PHYSIOLOGICAL AND BEHAVIORAL CHANGES IN
CC MATED FEMALE FLIES.
CC - SUBCELLULAR LOCATION: Secreted (Probable).
CC - SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U19729; AAB82348.1; -
DR COMPLEYEAST-2DPAGE; O13547; -
DR SGD; S0006429; CCW14.
KW Cell wall; Glycoprotein; GPI-anchor; Membrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 220
FT PROPEP 221 238
FT DOMAIN 77 218
FT CARBOHYD 87 87
FT LIPID 220 220
SQ SEQUENCE 238 AA; 23268 MW; 450042DC144C12B8 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 238;
Best Local Similarity 100.0%; Pred.No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 ASLGDS 14
Dbb 88 ASLGDS 93

RESULT 34
YCND_BACSU
ID YCND_BACSU STANDARD; PRT; 249 AA.
AC P94424;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical oxidoreductase ycmd (EC 1.-.-.-).
GN YCMD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97121189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
RT chromosome: determination of the sequence of a 146 kb segment and
RL identification of 113 genes."
RL Microbiology 142:3047-3056(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Betero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Braus A., Braun M., Brignell S.C., Bron S.,
RA Bouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Colightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haleb J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinol S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauei C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.J., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serroir P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,

```

Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler H., Weltzenegger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A., RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*." Nature 390:249-256(1997).  
RL Nature 390:249-256(1997).  
CC -!- COFACTOR: FMN (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE FLAVIN OXIDOREDUCTASE FRP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D50453; BAA09018.1; -.  
DR EMBL: Z99106; CAB12194.1; -.  
DR HSSP: Q56691; 1BKJ.  
DR SubtilList; BG12040; ycnD.  
DR InterPro: IPR000415; Nitroreductase.  
DR Pfam: PF00881; Nitroreductase; 1.  
DR Hypothetical protein; Oxidoreductase; Flavoprotein; FMN;  
KW Complete proteome.  
SQ SEQUENCE 249 AA; 27867 MW; 691BEEA4234FA59 CRC64;  
  
Query Match 6.2%; Score 6; DB 1; Length 249;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 25 KERKKK 30  
Db 54 KERKKK 59  
  
RESULT 35  
ID AGL6\_ARATH STANDARD; PRT; 252 AA.  
AC P29386;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Agamous-like MADS box protein AGL6.  
GN AGL6 OR AT2G43650 OR F17K2.18.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=91160981; PubMed=1672119;  
RX Ma H., Yanofsky M.F., Meyerowitz E.M.;  
RT "AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes." Genes Dev. 5:484-495(1991).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=cv. Columbia;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V., Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L., Morfitt K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L., Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana." Nature 402:761-768(1999).

CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION FACTORS.  
CC -!- SIMILARITY: CONTAINS 1 K-BOX DIMERIZATION DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M55554; AAA79328.1; -.  
DR EMBL: AC003680; AAC06173.1; -.  
DR PIR: F39534; F39534.  
DR HSSP: P11831; 1SRS.  
DR TRANSFAC; T03029; -.  
DR InterPro: IPR002487; TF\_Kbox.  
DR InterPro: IPR002100; TF\_MADSbox.  
DR Pfam: PF00319; SRP-TF; 1.  
DR Pfam: PF01486; K-box; 1.  
DR PRINTS; PR00404; MADSDOMAIN.  
DR SMART; SM00432; MADS; 1.  
DR PROSITE; PS00350; MADS\_BOX\_1; 1.  
DR PROSITE; PS00066; MADS\_BOX\_2; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein.  
FT DOMAIN 3 57 MADS.  
FT DOMAIN 95 167 K-BOX.  
SQ SEQUENCE 252 AA; 28744 MW; F763AA471515CF20 CRC64;  
  
Query Match 6.2%; Score 6; DB 1; Length 252;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 22 LRKKER 27  
Db 152 LRKKER 157  
  
RESULT 36  
YQEU\_BACSU STANDARD; PRT; 256 AA.  
ID YQEU\_BACSU  
AC P54461;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yqeu.  
GN YQEU.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=168 / JH642;  
RX MEDLINE=97124195; PubMed=8969508;  
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;  
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes." Microbiology 142:3103-3111(1996).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=168 / JH642;  
RX MEDLINE=97175542; PubMed=9023197;  
RA Homuth G., Masuda S., Mogk A., Kobayashi Y., Schumann W.;  
RT "The dnaK operon of *Bacillus subtilis* is heptacistronic." J. Bacteriol. 179:1153-1164(1997).  
RL [3]  
RN SEQUENCE FROM N.A.  
RP STRAIN=168;

```

RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertsch M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Chouillet S., Bruchli C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot K., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian F., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kleerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: BELONGS TO THE UPF0088 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D84432; BAA12467.1; -
CC DR EMBL; D83717; BAA12079.1; -
CC DR EMBL; Z99117; CAB14486.1; -
CC DR Subtilisin; BGL1645; Yqeu.
CC DR InterPro; IPR004382; Cons.hypoth46.
CC DR TIGRFAMs; TIGR00046; Cons.hypoth46; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 256 AA; 28802 MW; AAE708FE4283157E CRC64; .
CC -----
Query Match 6.2%; Score 6; DB 1; Length 256;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 KKKRR 33
Db 126 KKKRR 131
|||||
RESULT 37
FPG_VIBCH STANDARD; PRT; 269 AA.
AC Q9KVC5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) (Fapy-DNA
DE glycosylase).
GN MUTM OR FPG OR VC0221.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -!- FUNCTION: This enzyme may play a significant role in processes
CC leading to recovery from mutagenesis and/or cell death by
CC alkylating agents (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of DNA containing ring-opened N(7)-
CC methylguanine residues, releasing 2,6-diamino-4-hydroxy-5-(N-
CC methyl)formamidopyrimide.
CC -!- COFACTOR: Binds 1 zinc ion (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE004112; AAF93397.1; -
CC DR HSP; V05066; 1EE8.
CC DR TIGR; OC0221; -
CC DR InterPro; IPR000191; Fapy_DNA_glyco.
CC DR Pfam; PF01149; Fapy_DNA_glyco; 1.
CC DR ProDom; PD003680; Fapy_DNA_glyco; 1.
CC DR TIGRFAMs; TIGR00577; fpg; 1.
CC DR PROSITE; PS01242; FPG; FALSE-NEG.
CC KW DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;
CC Complete proteome.
CC FT ZN_FING 244 267 POTENTIAL.
CC SQ SEQUENCE 269 AA; 30036 MW; 677AV319E5F8887 CRC64; .
CC -----
Query Match 6.2%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVEVSR 6
Db 6 EVEVSR 11
|||||
RESULT 38
NRTB_PHOLA
ID NRTB_PHOLA STANDARD; PRT; 279 AA.
AC Q51881;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nitrate transport permease protein nrtb.
GN NRTB.
OS Phormidium laminosum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Phormidium.
OX NCBI_TaxID=32059;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=OH-1-P-CL1;
RA MEDLINE=95375238; PubMed=7647306;
RA Merchan F., Kindle K.L., Llana M.J., Serra J.L., Fernandez E.;
RA "Cloning and sequencing of the nitrate transport system from the
RA thermophilic, filamentous cyanobacterium Phormidium laminosum:
RA comparative analysis with the homologous system from Synechococcus

```

sp. PCC 7942.";  
 RL Plant Mol. Biol. 28:759-766(1995).  
 CC -!- FUNCTION: PROBABLY PART OF A HIGH-AFFINITY BINDING-PROTEIN-  
 CC DEPENDENT TRANSPORT SYSTEM FOR NITRATE; PROBABLY RESPONSIBLE FOR  
 CC THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT  
 CC SYSTEM PERMEASE FAMILY. CYSTW SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z19598; CAA79657.1; -  
 DR InterPro: IPR000515; BPD.transp.  
 DR Pfam: PF00528; BPD.transp; 1.  
 DR TIGRfams: TIGR01183; ntrb; 1.  
 DR PROSITE: PS00402; BPD\_TRANSF\_INN\_MEMBR; FALSE\_NEG.  
 DR TransPort: Transmembrane; Nitrate assimilation.  
 KW TRANSMEM 30 50 POTENTIAL.  
 FT TRANSMEM 98 118 POTENTIAL.  
 FT TRANSMEM 151 171 POTENTIAL.  
 FT TRANSMEM 196 216 POTENTIAL.  
 FT TRANSMEM 217 237 POTENTIAL.  
 FT TRANSMEM 249 269 POTENTIAL.  
 SQ SEQUENCE 279 AA; 30204 MW; 5BE86C4B02EFC5F6 CRC64;  
 -----  
 Query Match 6.2%; Score 6; DB 1; Length 279;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 QY 76 EMLTGG 81  
 Db 224 EMLTGG 229  
 -----  
 RESULT 39  
 NADCC\_METTJA STANDARD; PRT; 283 AA.  
 AC Q57916;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable nicotinate-nucleotide pyrophosphorylase [carboxylating]  
 DE (EC 2.4.2.19) (Quinolinate phosphoribosyltransferase  
 DE [decarboxylating]) (QAPRTase).  
 GN NADCC OR M0493.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kervatage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RA jannaschii."  
 RT Science 273:1058-1073(1996).  
 CC -!- CATALYTIC ACTIVITY: Nicotinate D-ribonucleotide + diphosphate +  
 CC CO(2) -> pyridine-2,3-dicarboxylate + 5-phospho-alpha-D-ribose 1-

CC diphosphate.  
 CC -!- PATHWAY: NAD biosynthesis; aspartate to NAMN; third (last) step.  
 CC -!- SIMILARITY: BELONGS TO THE NADCC/MODD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U67499; AAB98483.1; -  
 DR TIGR: M0493; -  
 DR InterPro: IPR004393; NADCC.  
 DR InterPro: IPR002638; QRPase.  
 DR Pfam: PF02749; QRPase; 1.  
 DR Pfam: PF02749; QRPase; 1.  
 DR ProDom: PD003988; QRPase; 1.  
 DR TIGRfams: TIGR00078; nadC; 1.  
 KW Pyridine nucleotide biosynthesis; Transferase; Glycosyltransferase;  
 KW Complete proteome.  
 SQ SEQUENCE 283 AA; 31970 MW; BDD118E9CE1401A8 CRC64;  
 -----  
 Query Match 6.2%; Score 6; DB 1; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 QY 39 CGIDFI 44  
 Db 48 CGIDFI 53  
 -----  
 RESULT 40  
 PARB\_CHLPN STANDARD; PRT; 286 AA.  
 AC Q9Z7M0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable chromosome partitioning protein parB.  
 GN PARB OR CPN0684 OR CP0062.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CHL029;  
 RX MEDLINE=99208606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RA "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RA pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138

RT from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- FUNCTION: INVOLVED IN CHROMOSOME PARTITION. LOCALIZE TO BOTH POLES  
 CC OF THE PREDIVISIONAL CELL FOLLOWING COMPLETION OF DNA REPLICATION.  
 CC BINDS TO THE DNA ORIGIN OF REPLICATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PARB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE001651; AAD18823.1; --  
 CC EMBL: AE002169; AAF37951.1; --  
 CC EMBL: AP002547; BAA98891.1; --  
 CC PHCI-2DPAGE; Q927M0; --  
 CC TIGR: CP0062; --  
 CC InterPro: IPR004437; ParB\_part.  
 CC InterPro: IPR003115; ParBc.  
 CC Pfam: PF02195; ParBc.  
 CC SMART: SM00470; ParBc; 1.  
 CC TIGRFAMs: TIGR00180; parB\_part; 1.  
 CC Chromosome partition; DNA-binding; Complete proteome.  
 CC KW SEQUENCE 286 AA; 32057 MW; 7E0C642FB82DA1B3 CRC64;  
 CC  
 CC Query Match 6.2%; Score 6; DB 1; Length 286;  
 CC Best Local Similarity 100.0%; Pred. No. 72;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 59 IQESLL 64  
 CC | | | | |  
 CC DB 158 IQESLL 163  
 CC  
 CC RESULT 41  
 CC ID EFTS\_ANASP STANDARD; PRT; 313 AA.  
 CC AC Q8MY3;  
 CC DT 15-JUN-2002 (Rel. 41, Created)  
 CC DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 CC DE Elongation factor Ts (EF-Ts).  
 CC GN TSF OR ALL4791.  
 CC OS Anabaena sp. (strain PCC 7120).  
 CC OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 CC OX NCBI\_TaxID=103690;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=21595285; PubMed=11759840;  
 CC RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 CC Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 CC Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 CC Nakazaki N., Shimpou S., Sugimoto M., Takazawa M., Yamada M.,  
 CC Yasuda M., Tabata S.;  
 CC "Complete genomic sequence of the filamentous nitrogen-fixing  
 CC cyanobacterium Anabaena sp. strain PCC 7120".  
 CC RL DNA Res. 8:205-213(2001).  
 CC -1- FUNCTION: Associates with the EF-Tu.GDP complex and induces the  
 CC exchange of GDP to GTP, it remains bound to the aminoacyl-tRNA.  
 CC EF-Tu.GTP complex up to the GTP hydrolysis stage on the ribosome.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE EF-TS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

CC EMBL: AP003597; BAB76490.1; --  
 CC InterPro: IPR001816; EF\_TS.  
 CC InterPro: IPR000449; UBA\_domain.  
 CC Pfam: PF00889; EF\_TS; 1.  
 CC Pfam: PF00627; UBA; 1.  
 CC TIGRFAMs: TIGR00116; tsf; 1.  
 CC PROSITE: PS01126; EF\_TS\_1; 1.  
 CC PROSITE: PS01127; EF\_TS\_2; 1.  
 CC KW Elongation factor; Protein biosynthesis; Complete proteome.  
 CC FT SITE 82 INVOLVED IN MG++ ION DISLOCATION FROM EF-  
 CC TU (BY SIMILARITY).  
 CC SQ SEQUENCE 313 AA; 34275 MW; 234F49543734652A CRC64;  
 CC  
 CC Query Match 6.2%; Score 6; DB 1; Length 313;  
 CC Best Local Similarity 100.0%; Pred. No. 78;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 EVEVSR 6  
 CC | | | | |  
 CC DB 280 EVEVSR 285  
 CC  
 CC RESULT 42  
 CC CPPI\_ENTHI STANDARD; PRT; 315 AA.  
 CC ID CPPI\_ENTHI  
 CC AC Q01957;  
 CC DT 01-JUN-1994 (Rel. 29, Created)  
 CC DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Cysteine proteinase 1 precursor (EC 3.4.22.-).  
 CC GN CPPI OR CPP.  
 CC OS Entamoeba histolytica.  
 CC OC Eukaryota; Entamoebidae; Entamoeba.  
 CC OX NCBI\_TaxID=5759;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 CC RC STRAIN=HM-1:IMSS;  
 CC RX MEDLINE=92389981; PubMed=1518524;  
 CC RA Tannich E., Nickel R., Buss H., Horstmann R.D.;  
 CC "Mapping and partial sequencing of the genes coding for two different  
 CC cysteine proteinases in pathogenic Entamoeba histolytica".  
 CC RL Mol. Biochem. Parasitol. 54:109-111(1992).  
 CC RN [2]  
 CC RP SEQUENCE OF 4-315 FROM N.A.  
 CC RC STRAIN=HM-1:IMSS;  
 CC RX MEDLINE=91161560; PubMed=1705935;  
 CC RA Tannich E., Scholze H., Nicke R., Horstmann R.D.;  
 CC "Homologous cysteine proteinases of pathogenic and nonpathogenic  
 CC Entamoeba histolytica. Differences in structure and expression.";  
 CC RL J. Biol. Chem. 266:4798-4803(1991).  
 CC RN [3]  
 CC RP SEQUENCE OF 94-99; 101-110 AND 112-113.  
 CC RC STRAIN=HM-1:IMSS;  
 CC RX MEDLINE=90095985; PubMed=2557443;  
 CC RA Schulte W., Scholze H.;  
 CC "Action of the major protease from Entamoeba histolytica on proteins  
 CC of the extracellular matrix".  
 CC RL J. Protozool. 36:538-543(1989).  
 CC -1- FUNCTION: INVOLVED IN THE DESTRUCTION OF HUMAN TISSUE BY  
 CC E.HISTOLYTICA. CAN ABOLISH ADHESION AND DEGRADE MATRIX PROTEINS  
 CC SUCH AS COLLAGEN, LAMININ AND FIBRONECTIN. MAY PLAY A ROLE  
 CC IMPORTANT ROLE IN PATHOGENICITY.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

```

DR EMBL; M94162; AAA29090.1; -.
DR EMBL; M64712; AAA29093.1; -.
DR HSSP; P07711; ICJL.
DR MEROPS; C01.050; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR PRODOM; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
DR PROSITE; PS00640; THIOLESTERASE_ASN; 1.
KW Hydrolase; Thiol protease; zymogen; Signal; Multigene family.
FT SIGNAL 1 13 POTENTIAL
FT PROPEP 14 93 ACTIVATION PEPTIDE.
FT CHAIN 94 315 CYSTEINE PROTEINASE 1.
FT ACT_SITE 118 118 BY SIMILARITY.
FT ACT_SITE 259 259 BY SIMILARITY.
FT ACT_SITE 279 279 BY SIMILARITY.
FT DISULFID 115 161 BY SIMILARITY.
FT DISULFID 152 193 BY SIMILARITY.
FT CONFLICT 4 4 F -> V (IN REF. 2).
FT CONFLICT 269 269 V -> A (IN REF. 2).
SQ SEQUENCE 315 AA; 35056 MW; DF4E3BC795164147 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 GDSSETL 17
DB 137 GDSSETL 142

RESULT 43
ID L767 CAEEL STANDARD; PRT; 316 AA.
AC Q09517;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative steroid dehydrogenase let-767 (BC 1.1.1.-).
GN LET-767 OR C56G2.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Connell M.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION
RA Kuevers L.M., O'Neil N.J., Baillie D.L.;
RT "let-767 is a gut-specific dehydrogenase.";
RL (In) Worm Breeder's Gazette 15(3):34(1998).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY. 17-BETA-HSD 3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U23177; AAA64333.2; -.

```

```

DR WormPep; C56G2.6; CE30639.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
FT NP_BIND 47 76 NADP (BY SIMILARITY).
FT ACT_SITE 202 202 BY SIMILARITY.
SQ SEQUENCE 316 AA; 34309 MW; DA3C6377AC4C12CE CRC64;

Query Match 6.2%; Score 6; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 LRKKER 27
DB 306 LRKKER 311

RESULT 44
ID H963 HUMAN STANDARD; PRT; 319 AA.
AC O14826;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable G protein-coupled receptor H963.
GN H963.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Peripheral blood monocytes;
RX MEDLINE=98036061; PubMed=9370294;
RA Jacobs K.A., Collins-Kacie L.A., Colbert M., Duckett M.,
RA Golden-Fleet M., Kelleher K., Kriz R., Lavallie E.R., Merberg D.,
RA Spaulding V., Stover J., Williamson M.J., McCoy J.M.;
RT "A genetic selection for isolating cDNAs encoding secreted proteins.";
RL Gene 198:289-296(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21426338; PubMed=11524702;
RA Joensuu T., Haemaelaenen R., Yuan B., Johnson C., Tegelberg S.,
RA Gasparini P., Zelante L., Pirvola U., Pakarinen L., Lehesjoki A.-E.,
RA de la Chapelle A., Sankila E.-M.;
RT "Mutations in a novel gene with transmembrane domains underlie Usher
RT syndrome type 3.";
RL Am. J. Hum. Genet. 69:673-684(2001).
CC -!- FUNCTION: ORPHAN RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF002986; AAC51846.1; -.
DR EMBL; AF411849; AAL47763.1; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCR_RHODOPSIN
DR PROSITE; PS00237; G_PROTEIN_RECEP_FL_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_FL_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 21 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 22 42 1 (POTENTIAL).
FT DOMAIN 43 43 CYTOPLASMIC (POTENTIAL).

```

QY 25 KERKKK 30  
Db 173 KERKKK 178

Search completed: May 11, 2003, 20:12:47  
Job time : 22.7434 secs

FT TRANSMEM 49 69 2 (POTENTIAL).  
FT DOMAIN 70 89 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 90 110 3 (POTENTIAL).  
FT DOMAIN 111 132 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 133 153 4 (POTENTIAL).  
FT DOMAIN 154 181 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 182 202 5 (POTENTIAL).  
FT DOMAIN 203 224 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 225 245 6 (POTENTIAL).  
FT DOMAIN 246 268 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 269 289 7 (POTENTIAL).  
FT DOMAIN 290 319 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 3 3 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 319 AA; 36754 MW; 79B9821C10841114 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 319;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTE 21  
Db 247 TLSQTE 252

RESULT 45  
PRIM\_BUCAP STANDARD; PRT; 319 AA.  
AC P32000;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA primase (EC 2.7.7.-) (Fragment).  
GN DNAG.  
OS Buchnera aphidicola (subsp. Schizaphis graminum).  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
OX NCBI\_TaxID=98794;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93012960; PubMed=1398077;  
RA Lal C.-F., Baumann P.;  
RT "Sequence analysis of a DNA fragment from Buchnera aphidicola (an endosymbiont of aphids) containing genes homologous to dnaG, rpoD, cysE, and secE.";  
RL Gene 119:113-118(1992).  
RC -!- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.  
CC -!- COPACITOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).  
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M90644; AAA73233.1; -  
DR PIR; PC1137; PC1137.  
DR HSSP; P02923; IEQN.  
DR InterPro; IPR002936; DNAPrim\_toprim.  
DR Pfam; PF01751; Toprim; 1.  
DR SMART; SM00493; TOPRIM; 1.  
KW Transferase; DNA replication; DNA-directed RNA polymerase; Primosome;  
KW Zinc-finger; Zinc; Metal-binding.  
FT NON\_TER 1  
SQ SEQUENCE 319 AA; 37483 MW; 868521F4AAF58341 CRC64;

Query Match 6.2%; Score 6; DB 1; Length 319;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2003, 19:33:41 ; Search time 51.5044 Seconds  
(without alignments)  
388.055 Million cell updates/sec

Title: US-09-854-133-586

Perfect score: 97

Sequence: 1 EVEYSRDHSLGDSSETLSQT.....LTGGCLPWATSHLGRKKS 97

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 65 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp Vertebrate.\*  
14: sp Unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7	7.2	36	2	O68941	O68941 rhodospiril
2	7	7.2	42	4	Q96RT9	Q96rt9 homo sapien
3	7	7.2	165	11	Q9CT47	Q9ct47 mus musculus
4	7	7.2	197	10	P93273	P93273 malus domes
5	7	7.2	203	10	Q9FTN4	Q9ftn4 oryza sativ
6	7	7.2	210	5	Q9VR27	Q9vr27 drosophila
7	7	7.2	235	17	Q975Z5	Q975z5 sulfolobus
8	7	7.2	248	10	Q948V1	Q948v1 magnolia pr
9	7	7.2	305	12	Q8QND6	Q8qnd6 ectocarpus
10	7	7.2	310	16	Q8ZPN9	Q8zpn9 salmonella
11	7	7.2	310	16	Q8Z6P0	Q8z6p0 salmonella
12	7	7.2	317	16	Q8R8B4	Q8r8b4 thermoanaer
13	7	7.2	318	10	Q9SL96	Q9sl96 arabidopsis
14	7	7.2	318	10	Q8VZ43	Q8vz43 arabidopsis
15	7	7.2	322	4	Q9HDA9	Q9hda9 homo sapien
16	7	7.2	340	17	Q9UYD9	Q9uyd9 pyrococcus

17	7	7.2	374	13	Q9DEV1
18	7	7.2	375	4	Q9HA26
19	7	7.2	383	16	Q9PPN7
20	7	7.2	394	10	Q9ZNU3
21	7	7.2	409	16	Q8YJDI
22	7	7.2	463	16	Q8YWS8
23	7	7.2	483	5	Q96148
24	7	7.2	483	8	Q36097
25	7	7.2	486	5	Q8WQA7
26	7	7.2	490	4	Q8WUZ1
27	7	7.2	517	4	Q76021
28	7	7.2	557	10	Q65538
29	7	7.2	560	16	Q86320
30	7	7.2	562	5	Q9NEY0
31	7	7.2	628	10	O81434
32	7	7.2	648	17	Q28883
33	7	7.2	651	10	Q9LNS7
34	7	7.2	680	5	Q9V6S8
35	7	7.2	699	5	O15816
36	7	7.2	703	16	Q9RZF3
37	7	7.2	772	4	Q9H1V1
38	7	7.2	777	10	O81868
39	7	7.2	799	4	Q96N31
40	7	7.2	881	5	Q9BKN8
41	7	7.2	1054	10	Q8RYX2
42	7	7.2	1283	5	Q95RH4
43	7	7.2	1286	5	O21025
44	7	7.2	1336	5	Q22944
45	7	7.2	1435	3	Q03291
46	7	7.2	1551	10	Q9SYPO
47	7	7.2	1911	5	Q9W4M7
48	7	7.2	1987	5	Q9V841
49	7	7.2	2434	8	Q9WTH5
50	7	7.2	3844	5	Q34648
51	6	6.2	50	16	Q8VKK7
52	6	6.2	63	17	O29078
53	6	6.2	74	16	Q8YBH2
54	6	6.2	80	6	Q9GMW9
55	6	6.2	80	10	Q9SJ65
56	6	6.2	89	16	Q9K5X5
57	6	6.2	90	5	Q9NLX9
58	6	6.2	103	12	O89411
59	6	6.2	104	4	Q9POE3
60	6	6.2	106	16	Q98A87
61	6	6.2	110	10	O64396
62	6	6.2	110	17	Q8ZZZ4
63	6	6.2	111	11	Q9QVN7
64	6	6.2	116	12	O8V6V6
65	6	6.2	116	16	Q9WYJ5

## ALIGNMENTS

RESULT 1

O68941	PRELIMINARY;	PRT;	36 AA.
ID	O68941		
AC	O68941		
DT	01-AUG-1998 (TREMBLrel. 07, Created)		
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Dinitrogenase 3 beta subunit (Fragment).		
GN	ANFK.		
OS	Rhodospirillum rubrum.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;		
OC	Rhodospirillum.		
OX	NCBI_TaxID=1085;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Loveless T.M., Bishop P.E.;		
RT	"Identification of Genes Unique to Mo-Independent Nitrogenase Systems in Diverse Diazotrophs.";		
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.		

Q9dev1 brachydanio  
Q9ha26 homo sapien  
Q9ppn7 ureaplasma  
Q9znv3 arabidopsis  
Q8yjd1 bruceella me  
Q8yw58 anabaena sp  
Q96148 plasmodium  
Q96097 theileria p  
Q8wqa7 caenorhabdi  
Q8wuz1 homo sapien  
Q76021 homo sapien  
Q65538 arabidopsis  
Q86320 mycobacteri  
Q9ney0 caenorhabdi  
O81434 arabidopsis  
O28883 archaeoglob  
Q9lns7 arabidopsi  
Q9v6s8 drosophila  
O15816 dictyostell  
Q9zrf3 deinococcus  
Q9h1v1 homo sapien  
O81868 arabidopsis  
Q96n31 homo sapien  
Q9bkn8 plasmodium  
Q95rh4 drosophila  
Q21025 caenorhabdi  
Q22944 caenorhabdi  
Q03291 saccharomyc  
Q9sypp0 arabidopsi  
Q9w4m7 drosophila  
Q9v841 drosophila  
Q9wth5 oenothera h  
Q94648 plasmodium  
Q8vkk7 mycobacteri  
O29078 archaeoglob  
Q8ybh2 bruceella me  
Q9gmw9 macaca fasc  
Q9s165 arabidopsis  
Q9nlx9 leishmania  
Q89411 paramecium  
Q9poe3 homo sapien  
Q98a87 rhizobium l  
O64396 pisum sativ  
Q8zxz4 pyrobaculum  
Q9qvn7 mus sp. sili  
Q8v6v6 halovirus h  
Q9wyj5 thermotoga

```

DR EMBL; AF058778; AAC14327.1; -.
DR InterPro; IPR000510; Oxred_nitroonsel.
DR Pfam; PF00148; oxidored_nitro; 1.
DR NON_TER 36
SQ SEQUENCE 36 AA; 3957 MW; D9AF46BCFD437D97 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKKR 27
Db 4 ELRKKR 10

RESULT 2
Q96RT9 ID Q96RT9 PRELIMINARY; PRT; 42 AA.
AC Q96RT9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HBV pRESI(21-47) binding protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang J.-Y., Wang Y.;
RT "Cloning and identification of a novel pRESI(21-47) binding protein,
RT SIBP."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF271256; AAK58397.1; -.
FT NON_TER 1
SQ SEQUENCE 42 AA; 5559 MW; 52CDC6CAADF9C2FB CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKRRK 29
Db 3 RKKRRK 9

RESULT 3
Q9CT47 ID Q9CT47 PRELIMINARY; PRT; 165 AA.
AC Q9CT47;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DNA polymerase epsilon, subunit 2 (Fragment).
GN POLE2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staehli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

```

```

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK011194; BAB27458.1; -.
DR MGD; MGI:1197514; Pole2.
FT NON_TER 165
SQ SEQUENCE 165 AA; 19229 MW; 15933826B119C38D CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 RKKRRK 29
Db 158 RKKRRK 164

RESULT 4
P93273 ID P93273 PRELIMINARY; PRT; 197 AA.
AC P93273;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PAFD103 protein (Fragment).
GN PAFD103.
OS Malus domestica (Apple) (Malus sylvestris).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Rosales; Rosaceae; Maloideae; Malus.
OX NCBI_TaxID=3750;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GRANNY SMITH;
RA Dong Y.-H., Janssen B.J., Bielecki L.L., Atkinson R.G., Morris B.A.,
RA Gardner R.C.;
RT "Isolating and characterizing genes differentially expressed early in
RT apple fruit development."
RL J. Am. Soc. Hortic. Sci. 122:752-757(1997).
DR EMBL; U80270; AAC06385.1; -.
FT NON_TER 1
SQ SEQUENCE 197 AA; 21714 MW; 92518666575C3690 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKKRRK 30
Db 54 KKKRRK 60

RESULT 5
Q9FTN4 ID Q9FTN4 PRELIMINARY; PRT; 203 AA.
AC Q9FTN4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE P0005A05.15 protein.
GN P0005A05.15.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;

```

```

[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:PO005A05.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002863; BAB16911.1; -.
SQ SEQUENCE 203 AA; 22741 MW; 1A74C919A6F70852 CRC64;

Query Match
Best Local Similarity 7.2%; Score 7; DB 10; Length 203;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 RKKRER 33
DB 186 RKKRER 192
|||||

RESULT 6
Q9VR27 PRELIMINARY; PRT; 210 AA.
AC Q9VR27;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG3244 protein.
GN CG3244.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., DeLcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```

```

"The genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195(2000).
RL EMBL; AE003575; AAF50981.2; -.
DR FlyBase; FBgn0031629; CG3244.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C.TYPE.LECTIN.1; UNKNOWN.1.
DR PROSITE; PS00041; C.TYPE.LECTIN.2; 1.
DR PROSITE; PS00041; C.TYPE.LECTIN.2; 1.
SQ SEQUENCE 210 AA; 24368 MW; DA0D011A96D64430 CRC64;

Query Match
Best Local Similarity 7.2%; Score 7; DB 5; Length 210;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 CPPSPKE 71
DB 21 CPPSPKE 27
|||||

RESULT 7
Q975Z5 PRELIMINARY; PRT; 235 AA.
AC Q975Z5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein ST0283.
GN ST0283.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudo H., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000982; BAB5253.1; -.
DR InterPro; IPR000836; PRtransferase.
DR Pfam; PF00156; Ribosyltran; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 235 AA; 26863 MW; 0F82B46357A350C4 CRC64;

Query Match
Best Local Similarity 7.2%; Score 7; DB 17; Length 235;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 IQESLLC 65
DB 35 IQESLLC 41
|||||

RESULT 8
Q948V1 PRELIMINARY; PRT; 248 AA.
AC Q948V1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative MADS-domain transcription factor MpMADS4 (Fragment).
GN MpMADS4.
OS Magnolia praecocissima.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Magnoliales; Magnoliaceae; Magnolia.

```

```

OX NCBI_TaxID=81865;
RN [1]
RP SEQUENCE FROM N.A.
RA Ito M., Shiobara S., Tanabe Y., Hasebe M.;
RT "Organ identities in Magnolian flower.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB050646; BAB70739.1; -.
DR InterPro: IPR002487; TF_Rbox.
DR Pfam: PF01486; K-box; 1.
DR PFam: PF00319; SRF-TF; 1.
DR PROSITE: PS50066; MADS_BOX_2; 1.
FT NON_TER
SQ SEQUENCE 248 AA; 28206 MW; A84E80942F0B6D58 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 248;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKKR 27
DB 136 ELRKKR 142

RESULT 9
Q8QND6 PRELIMINARY; PRT; 305 AA.
ID Q8QND6;
AC Q8QND6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Esv-1-146.
OS Ectocarpus siliculosus virus.
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phaeovirus.
OX NCBI_TaxID=37665;
RN [1]
RP SEQUENCE FROM N.A.
RA Delarouque N., Bothe G., Pohl T., Knippers R., Mueller D.G., Boland W.;
RT "The complete nucleotide sequence of the Ectocarpus siliculosus virus genome.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF204951; AAK14561.1; -.
SQ SEQUENCE 305 AA; 33947 MW; 1B86898C0CDB999 CRC64;

Query Match 7.2%; Score 7; DB 12; Length 305;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 REMLTGG 81
DB 141 REMLTGG 147

RESULT 10
Q8ZPN9 PRELIMINARY; PRT; 310 AA.
ID Q8ZPN9;
AC Q8ZPN9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transcriptional regulator, Lysr family.
GN YDHB OR STM1429.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

```

```

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
RL Nature 413:852-856(2001).
DR EMBL: AE008762; AAL20351.1; -.
DR InterPro: IPR000847; HTH_Lysr.
DR Pfam: PF00126; HTH_1; 1.
DR Pfam: PF03466; Lysr_subst.
DR PRINTS: PR00039; HTHLYSR.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 310 AA; 35013 MW; 7C75EBAFBA2C630 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 310;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17
DB 288 LGDSETL 294

RESULT 11
Q8Z6P0 PRELIMINARY; PRT; 310 AA.
ID Q8Z6P0;
AC Q8Z6P0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transcriptional regulator.
GN Sty1693.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL: AL627271; CAD01938.1; -.
DR InterPro: IPR000847; HTH_Lysr.
DR Pfam: PF00126; HTH_1; 1.
DR Pfam: PF03466; Lysr_subst.
DR PRINTS: PR00039; HTHLYSR.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 310 AA; 35058 MW; 41C8CC313DA26E8A CRC64;

Query Match 7.2%; Score 7; DB 16; Length 310;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LGDSETL 17
DB 288 LGDSETL 294

RESULT 12

```

```

Q8R8B4
ID Q8R8B4 PRELIMINARY; PRT; 317 AA.
AC Q8R8B4
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein TTE2096.
GN TTE2096.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao O., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AF013157; AAM25267.1; -.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 317 AA; 35043 MW; FB22219729DEF996 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 317;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 IFWILLF 54
DB 16 IFWILLF 22
|||||||

RESULT 13
Q9SL96 PRELIMINARY; PRT; 318 AA.
ID Q9SL96
AC Q9SL96
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE AT2925670 protein.
GN AT2925670.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gili J.E.,
RA Adams M.D., Carrera A.J., Cressy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
RL thaliana.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX Lin X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006053; AAD31367.1; -.
SQ SEQUENCE 318 AA; 35128 MW; 2F2E2CD269FFD004 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 318;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 24 KKERKKK 30
DB 160 KKERKKK 166
|||||||

RESULT 14
Q8VZ43 PRELIMINARY; PRT; 318 AA.
ID Q8VZ43
AC Q8VZ43
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical 35.2 kDa protein.
GN AT2G25670.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC Yamada K., Ban J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
RC Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene At2g25670 (GI:15225169).";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY065267; AAL38743.1; -.
KW Hypothetical protein.
SQ SEQUENCE 318 AA; 35154 MW; 2F2E2CC278FED004 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 318;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKKK 30
DB 160 KKERKKK 166
|||||||

RESULT 15
Q9HDA9 PRELIMINARY; PRT; 322 AA.
ID Q9HDA9
AC Q9HDA9
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 36.7 kDa protein (Fragment).
GN CAPX-11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kairo A., Wang L., Gao Z.Q., Gao Z.P., Roman B.M.;
RT "Isolation of Novel Genes from Human Colonic Epithelial Cells.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083127; AAF98239.1; -.
DR InterPro; IPR002143; Ribosomal_L1.
DR Pfam; PF00687; Ribosomal_L1; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 322 AA; 36676 MW; E2B78F862E7B52B CRC64;

Query Match 7.2%; Score 7; DB 4; Length 322;
Best Local Similarity 100.0%; Pred. No. 26;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31  
 Db 225 KERKKR 231

## RESULT 16

Q9UYD9 PRELIMINARY; PRT; 340 AA.  
 AC Q9UYD9;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein PAB1036.  
 GN PAB1036.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ORSAY;  
 RA Hellig R.;  
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ248288; CAB50473.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 340 AA; 40246 MW; 30781FCE6B59C3C5 CRC64;

Query Match 7.2%; Score 7; DB 17; Length 340;

Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31  
 Db 287 KERKKR 293

## RESULT 17

Q9DEV1 PRELIMINARY; PRT; 374 AA.  
 AC Q9DEV1;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Lunatic fringe precursor.  
 GN LFNG.  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Leve C., Gajewski M., Tautz D.;  
 RT "Lunatic fringe mRNA."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY007434; AAG12160.1; -;  
 DR ZFIN; ZDB-GENE-960605-16; lfng.  
 DR InterPro; IPR003378; Fringe.  
 DR Pfam; PF02434; Fringe; 1.  
 KW Signal.  
 FT CHAIN 1 35 POTENTIAL.  
 FT CHAIN 84 374 LUNATIC FRINGE.  
 SQ SEQUENCE 374 AA; 41881 MW; 082F1FD0705B9A8B CRC64;

Query Match 7.2%; Score 7; DB 13; Length 374;

Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SLGDSSET 16

Db 51 SLGDSSET 57

## RESULT 18

Q9HA26 PRELIMINARY; PRT; 375 AA.  
 AC Q9HA26;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE CDNA FLJ12363 fis, clone MAMMAL002380.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MAMMARY GLAND;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,  
 RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Takiguchi S.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
 RA Ninomiya K., Iwayanagi T.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK022425; BAB14033.1; -;  
 DR InterPro; IPR004012; Run.  
 DR Pfam; PF02759; RUN; 1.  
 SQ SEQUENCE 375 AA; 41861 MW; F7DFCB4297CD95 CRC64;

Query Match 7.2%; Score 7; DB 4; Length 375;

Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KERKKK 30  
 Db 254 KERKKK 260

## RESULT 19

Q9PPN7 PRELIMINARY; PRT; 383 AA.  
 AC Q9PPN7;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical membrane lipoprotein.  
 GN UUG02.  
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 OC Mycoplasmataceae; Ureaplasma.  
 OX NCBI\_TaxID=134821;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=SEROVAR 3;  
 RX MEDLINE=20500219; PubMed=11048724;  
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
 RA Cassell G.H.;  
 RT "The complete sequence of the mucosal pathogen Ureaplasma urealyticum."  
 RL Nature 407:757-762(2000).  
 DR EMBL; AE002158; AAF31016.1; -;  
 DR InterPro; IPR001708; 60kDa\_innerneb.  
 DR Pfam; PF02096; 60KD\_IMP; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 383 AA; 43714 MW; DD20085C25CDE265 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 383;

Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 24 KKERKK 30
    |||||
DB 30 KKERKK 36

RESULT 20
Q92NU3
ID Q92NU3 PRELIMINARY; PRT; 394 AA.
AC Q92NU3
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative extensin.
GN AT2G46630.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC005819; AAC69930.1; -.
RL EMBL; AC006418; AAM15241.1; -.
DR InterPro; IPR000412; ABCtransprtr2.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
DR PROSITE; PS00890; ABC2.MEMBRANE; UNKNOWN.1.
SQ SEQUENCE 394 AA; 43246 MW; 55364D8A9DE3631 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 394;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 66 PPSPEV 72
    |||||
DB 124 PPSPEV 130

RESULT 21
Q8YJDI
ID Q8YJDI PRELIMINARY; PRT; 409 AA.
AC Q8YJDI
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative allantoin permease.
GN BMEI0155.

Query Match 7.2%; Score 7; DB 10; Length 394;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 KKERKK 30
    |||||
DB 30 KKERKK 36

RESULT 20
Q92NU3
ID Q92NU3 PRELIMINARY; PRT; 394 AA.
AC Q92NU3
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative extensin.
GN AT2G46630.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC005819; AAC69930.1; -.
RL EMBL; AC006418; AAM15241.1; -.
DR InterPro; IPR000412; ABCtransprtr2.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
DR PROSITE; PS00890; ABC2.MEMBRANE; UNKNOWN.1.
SQ SEQUENCE 394 AA; 43246 MW; 55364D8A9DE3631 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 394;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 66 PPSPEV 72
    |||||
DB 124 PPSPEV 130

RESULT 21
Q8YJDI
ID Q8YJDI PRELIMINARY; PRT; 409 AA.
AC Q8YJDI
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative allantoin permease.
GN BMEI0155.

Query Match 7.2%; Score 7; DB 10; Length 394;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 ELKKER 27
    |||||
DB 199 ELKKER 205

RESULT 23

```

```

OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIONTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Muejer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009458; AAL51337.1; -.
DR InterPro; IPR001248; Cyt_pur_permease.
DR Pfam; PF02133; Transp_cyt_pur; 1.
KW Complete proteome.
SQ SEQUENCE 409 AA; 44996 MW; BDIAC6559DD768B9 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 409;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 SQTELK 24
    |||||
DB 396 SQTELK 402

RESULT 22
Q8YW58
ID Q8YW58 PRELIMINARY; PRT; 463 AA.
AC Q8YW58
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein A111758.
GN A111758.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003587; BAB73457.1; -.
DR InterPro; IPR003018; GAF.
DR InterPro; IPR001932; PP2C-like.
DR Pfam; PF01590; GAF; 1.
DR SMART; SM00065; GAF; 1.
DR SMART; SM00331; PP2C_SIG; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 463 AA; 52033 MW; E6BCADFDEA8B7AB1 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 463;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 ELKKER 27
    |||||
DB 199 ELKKER 205

RESULT 23

```

```

096148 ID 096148 PRELIMINARY; PRT; 483 AA.
AC 096148;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DE 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE Hypothetical 57.8 kDa protein.
GN PFE0235W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetteilin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL: AF001382; AAC71836.1; -.
KW Hypothetical protein.
SQ SEQUENCE 483 AA; 57785 MW; 9332C1EC9B4DECFB CRC64;

Query Match 7.2%; Score 7; DB 5; Length 483;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30
Db 459 KKERKK 465

RESULT 24
Q36097 ID Q36097 PRELIMINARY; PRT; 483 AA.
AC Q36097;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).
GN Col.
OS Theileria parva.
OG Mitochondrion.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC Theileria.
OX NCBI_TaxID=5875;
RN [1]
RX STRAIN=MUGUGA;
RX MEDLINE=94155854; PubMed=8112303;
RA Kairo A., Fairlamb A., Gobright E., Nene V.;
RT "A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA
RT sequences and open reading frames for mitochondrially-encoded
RT proteins.";
RL EMBO J. 13:898-905(1994).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 4 FERROCYTOCHROME
CC C + 2 H(2)O.
CC -1- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC EMBL: 223263; CAA80798.1; -.

```

```

DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 483 AA; 54008 MW; 314438D6EF4CF3D6 CRC64;

Query Match 7.2%; Score 7; DB 8; Length 483;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 GIDFIIF 46
Db 150 GIDFIIF 156

RESULT 25
Q8WQ7 ID Q8WQ7 PRELIMINARY; PRT; 486 AA.
AC Q8WQ7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Y10588A.17 protein.
GN Y10588A.17
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RX SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases..
RN [2]
RX SEQUENCE FROM N.A.
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: AL132876; CAD21666.1; -.
SQ SEQUENCE 486 AA; 55840 MW; A2C9EA88AFA79D0E CRC64;

Query Match 7.2%; Score 7; DB 5; Length 486;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKKER 27
Db 235 ELRKKER 241

RESULT 26
Q8WU21 ID Q8WU21 PRELIMINARY; PRT; 490 AA.
AC Q8WU21;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 55.0 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC019069; AAH19069.1; -.
DR InterPro: IPR000637; AT_hook.
DR InterPro: IPR002143; Ribosomal_L1.
DR Pfam: PF00687; Ribosomal_L1; 1.

```



DR PRINTS; PR00929; ATHOOK.  
KW Hypothetical protein.  
SQ SEQUENCE 490 AA; 54972 MW; 55CDB8AA8BC3709 CRC64;

Query Match 7.2%; Score 7; DB 4; Length 490;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31  
| | | | |  
Db 300 KERKKR 306

## RESULT 27

O76021 PRELIMINARY; PRT; 517 AA.  
AC O76021;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE PBK1 protein.  
GN PBK1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=99075201; PubMed=9859858;  
RA Huch G., Hohn H.P., Denker H.W.;  
RT "Identification of differentially expressed genes in human trophoblast cells by DDRT-PCR";  
RL Placenta 19:557-567(1998);  
DR EMBL; AJ007398; CAA07491.1; -;  
DR SWISS-2DPAGE; O76021; HUMAN.  
DR InterPro; IPR002143; Ribosomal\_L1.  
DR Pfam; PF00687; Ribosomal\_L1; 2.  
SQ SEQUENCE 517 AA; 58132 MW; BB16E3D96A5EE27A CRC64;

Query Match 7.2%; Score 7; DB 4; Length 517;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KERKKR 31  
| | | | |  
Db 300 KERKKR 306

## RESULT 28

O65538 PRELIMINARY; PRT; 557 AA.  
AC O65538;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Hypothetical 61.6 kDa protein.  
GN F4D1.190 OR AT4G32610.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Hoheisel J.,  
RA Meves H.W., Mayer K.F.X., Schueller C.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Beves V., Rechmann S., Borkova D., Ansoerge W., Meves H.W., Lemcke K.,  
RA Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL022537; CAA18600.1; -;  
DR EMBL; AL161581; CAB79978.1; -;  
DR InterPro; IPR003428; MAM33.  
DR Pfam; PF02330; MAM33; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 557 AA; 61561 MW; 90CDA4D869601609 CRC64;

Query Match 7.2%; Score 7; DB 10; Length 557;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 KKERKK 30  
| | | | |  
Db 163 KKERKK 169

## RESULT 29

O86320 PRELIMINARY; PRT; 560 AA.  
AC O86320;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein Rv0976c.  
GN Rv0976c OR MTV044.04C OR MT1003.1.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RA Parkhill J.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / OSHKOSH;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., S.L.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Kacher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z79700; CAB02013.1; -;  
DR EMBL; AF006984; AAK45253.1; -;  
DR TIGR; MT1003; -;  
DR TubercuList; Rv0976c; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 560 AA; 59122 MW; 1DBD773C4814AE90 CRC64;

Query Match 7.2%; Score 7; DB 16; Length 560;  
Best Local Similarity 100.0%; Pred. No. 41;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 REMLTGG 81  
|||||  
Db 18 REMLTGG 24

RESULT 30

Q9NEY0  
ID Q9NEY0 PRELIMINARY; PRT; 562 AA.

DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Y105E8E.t protein.  
GN Y105E8E.t

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID-6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Sulston J.E.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE-99069613; PubMed-9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology";

RL Science 282:2012-2018(1998).

DR EMBL; AL132880; CAB60883.2; -

DR InterPro; IPR001005; Myb\_DNA\_binding.

DR SMART; SM00395; SANT; 1.

SQ SEQUENCE 562 AA; 64912 MW; FDA433D3D888DEF2 CRC64;

Query Match

Best Local Similarity 7.2%; Score 7; DB 5; Length 562;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ELRKKR 27

|||||

Db 235 ELRKKR 241

RESULT 31

O81434

ID O81434

AC O81434;

DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE T24H24.5 protein.

GN T24H24.5 OR AT4G04070.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI\_TaxID-3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RA WASHU;

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RA Courtney L., Stoneking T., Langston Y., Mead K.;

RL "The sequence of A. thaliana T24H24.";

RN Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases..

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RA Waterston R.;

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,

RA Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF075598; AAC28203.1; -

DR EMBL; AL161499; CAB77875.1; -

KW Hypothetical protein.

SQ SEQUENCE 628 AA; 68791 MW; B09D244B764722EB CRC64;

Query Match

Best Local Similarity 7.2%; Score 7; DB 10; Length 628;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TLSQTEL 22

|||||

Db 222 TLSQTEL 228

RESULT 32

O28883

ID O28883

AC O28883;

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE DNA helicase, putative.

GN AF1388.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

OC Archaeoglobaceae; Archaeoglobus.

OX NCBI\_TaxID-2234;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE-98049343; PubMed-9389475;

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

RA Richardson D.L., Kerlavage A.R., Gwinn M., Hickey E.K., Peterson J.D.,

RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,

RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glöckle A., Zhou L.,

RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,

RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,

RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,

RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,

RA Venter J.C.;

RL "The complete genome sequence of the hyperthermophilic, sulphate-

reducing archaeon Archaeoglobus fulgidus.";

RL Nature 390:364-370(1997).

DR EMBL; AE001009; AAB89860.1; -

DR TIGR; AF1388; -

DR InterPro; IPR004483; put\_DNA\_helic.

DR TIGRFAMs; TIGR00376; put\_DNA\_helic; 1.

KW Hypothetical protein; Helicase; Complete proteome.

SQ SEQUENCE 648 AA; 73775 MW; B5074507D856A7D5 CRC64;

Query Match

Best Local Similarity 7.2%; Score 7; DB 17; Length 648;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 GDSFELS 18

|||||

Db 609 GDSFELS 615